DOCUMENT RESUME

ED 194 590 TM 800 720

AUTHOR	Johnson, Dale D.: And Others
TITLE	The Word Identification Test Battery: A New Approach to Mastery and the Assessment of Word Identification Skills. Report from the Project on Studies in Language: Reading and Communication.
INSTITUTION	Wisconsin Univ., Madison. Research and Development Center for Individualized Schooling.
SPONS AGENCY	National Inst. of Education (DHEW), Washington, D.C.
REPORT NO	WRDCIS-TR-553
PUB DATE	Sep 80
GRANT	OB-NIE-G-8-0117
NOTE	178p.; For a related document see TM 800 704.
EDRS PRICE	MF01/PC08 Plus Postage.
DESCRIPTORS	Elementary Education: Language Skills: Mastery Learning: *Mastery Tests: *Phonics: *Structural Analysis (Linguistics): *Test Construction: *Word Recognition
IDENTIFIERS	*Word Identification Test (Johnson et al)

ABSTRACT

The work reported culminates research by the Project on the Assessment and Analysis of Word Identification Skills in Reading. The Word Identification Test battery was designed for elementary school children, with attention to the major issues pertaining to skills mastery and assessment that are raised in the review of mastery learning. Five important areas were of concern in the development of the battery: (a) basis on which target skills would be selected for inclusion: (b) facilitation of error analysis by creating categorical distractors: (c) ease and efficiency of test administration: (d) independence of the test battery from published materals to lessen the likelihood of teachers teaching to the tests: and (e) establishment of flexible standards for skills mastery based on a global measure of comprehension, rather than on arbitrary cutoff scores. The battery is comprised of five subtests within two major skill areas, phonics and structural analysis. The battery is a valid reliable instrument, and is easy to administer. It can facilitate diagnostic decisions about apportionment of instructional time on the most frequently occurring phonics and structural elements. Performance standards are provided for each subtest. (Author/GK)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

SCOPE OF INTEREST NOTICE

The ERIC Facility has assigned this document for processing to:

TM | CS

In our judgement, this document is also of interest to the clearinghouses noted to the right. Indexing should reflect their special points of view.

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

[illegible]

by Dale D. Johnson, Susan D. Pittman, Lisa K. Smberg, Judy Schwenker, and Barbara S. Dahl

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

R. Rossmiller

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Technical Report No. 553

THE WORD IDENTIFICATION TEST BATTERY: A NEW APPROACH TO MASTERY AND THE ASSESSMENT OF WORD IDENTIFICATION SKILLS

by

Dale D. Johnson, Susan D. Pittelman,
Linda K. Shriberg, Judy Schwenker, and Sandra S. Dahl

Report from the Project on
Studies in Language: Reading and Communication

Dale D. Johnson
Faculty Associate

Wisconsin Research and Development Center
for Individualized Schooling
The University of Wisconsin
Madison, Wisconsin

September 1980

Published by the Wisconsin Research and Development Center for Individualized Schooling. The project presented or reported herein was performed pursuant to a grant from the National Institute of Education, Department of Health, Education, and Welfare. However, the opinions expressed herein do not necessarily reflect the position or policy of the National Institute of Education, and no official endorsement by the National Institute of Education should be inferred.

Center Grant No. OB-NIE-G-80-0117

MISSION STATEMENT

The mission of the Wisconsin Research and Development Center is to improve the quality of education by addressing the full range of issues and problems related to individualized schooling. Teaching, learning, and the problems of individualization are given concurrent attention in the Center's efforts to discover processes and develop strategies and materials for use in the schools. The Center pursues its mission by

- conducting and synthesizing research to clarify the processes of school-age children's learning and development
- conducting and synthesizing research to clarify effective approaches to teaching students basic skills and concepts
- developing and demonstrating improved instructional strategies, processes, and materials for students, teachers, and school administrators
- providing assistance to educators which helps transfer the outcomes of research and development to improved practice in local schools and teacher education institutions

The Wisconsin Research and Development Center is supported with funds from the National Institute of Education and the University of Wisconsin.

WISCONSIN RESEARCH AND DEVELOPMENT
CENTER FOR INDIVIDUALIZED SCHOOLING

Acknowledgments

The authors gratefully acknowledge the cooperation of the pupils and teachers at Glendale, Huegel, and Thoreau Elementary Schools in Madison, Wisconsin; Northside, Sauk Trail, and West Middleton Elementary Schools in Middleton, Wisconsin; Oakland Elementary School in Antioch, Illinois; Holman Area Elementary School in Holman, Wisconsin; Waterloo Elementary School in Waterloo, Wisconsin; Keshena and Neopit Elementary Schools in Menominee, Wisconsin; Fairview North and Swanson Elementary Schools in Brookfield, Wisconsin; Algonquin and Dean Elementary Schools and Brown Deer Middle School in Brown Deer, Wisconsin; Kegonsa Elementary School and Stoughton Middle School in Stoughton, Wisconsin; Kado Road Elementary School in North Little Rock, Arkansas; Atkinson Elementary School in Portland, Oregon; Wilson Elementary School in San Gabriel, California; Kennelworth Elementary School in Phoenix, Arizona; Oregon Elementary School and Oregon Middle School in Oregon, Wisconsin; Marshall Elementary School in Marshall, Wisconsin; Eagle River Elementary School in Eagle River, Wisconsin; Antigo Elementary School in Antigo, Wisconsin and Belleville Elementary School in Belleville, Wisconsin. Their support of and participation in this research is greatly appreciated.

The authors also acknowledge with gratitude Lawrence D. Shriberg, Porfessor of Speech Pathology, University of Wisconsin-Madison; Dominic W. Massaro, Professor of Psychology, University of Wisconsin-Madison; Peter A. Schreiber, Professor of English and Linguistics, University of Wisconsin-Madison; and Richard L. Venezky, Unidel Professor of Educational Foundations, University of Delaware, for their helpful suggestions.

The authors also wish to acknowledge Judy Ross, Evelyn Read, and Judy Rosberg for their assistance with data collection. A special thank you is certainly owed to Carol Morgan-Janty and Mary Murphy for their valuable contributions to these studies.

Table of Contents

Page

Acknowledgments iv

List of Tables xi

List of Figures. xiii

Abstract xv

Introduction 1

A Review of the Instructional Trends and Assessment Instruments in the Areas of Phonics and Structure. 7

Phonics 7
Issues of Test Construction. 11
Scope of Test 11
Format of Test. 12
Presentation of Target Spelling-to-Sound Correspondences and Response Choices. 14
The Phonics Component of the Word Identification Test Battery 18
Structural Analysis 19
Derivatives. 21
Variants 23
Compound Words 23
Structure Component of the Word Identification Test Battery 25
Inflected Endings Subtest 27
Affixes Subtest 29
Contractions & Possessives Subtest. 31
Summary 33

Introduction to Mastery Learning 35

An Historical Perspective 35
Competency-based Education. 38
Definition of Mastery Learning. 39
Carroll's Model of School Learning 39
Bloom's Theory of School Learning. 41
Learning Tasks and Units. 43
Cognitive Entry Behaviors 43
Affective Entry Behaviors 44
Quality of Instruction. 44
Affective Outcomes. 45
Rate of Learning. 45
Level and Type of Achievement 46

Page

Mastery Learning in the Classroom 47
Two Basic Instructional Designs. 47
Bloom's Learning for Mastery Program. 47
Keller's Personalized System of Instruction 48
Noted Effects of Mastery Learning Theory 49
Research on the Components of Mastery Learning Models . 49
Cognitive Achievement. 49
Retention of Learning. 50
Transfer of Learning 50
Affective Characteristics. 51
Rate of Learning 52
Additional Current Research Summaries. 52
Mastery Learning Theory: Some Applications
to Reading Instruction 55
Mastery Learning and Reading Instruction 56
Objective-Based Reading Instruction. 58
Minimal Competencies in Reading. 59
The Measurement and Evaluation of Mastery Learning. 61
Determining the Appropriate Type of Assessment Instrument. 62
Norm-referenced Tests 62
Definition and Uses of Criterion-referenced Tests (CRT) 63
The Development of Criterion-referenced Tests. . . . 63
The Issue of Performance Standards 65

Final Evaluation of the Word Identification Test Battery 71

Method. 71
Subjects 71
Stimuli. 73
Phonics Subtests. 73
Consonants Subtests. 73
Vowels Subtest 78
Structure Subtests. 83
Inflected Endings Subtest. 83
Contractions & Possessives Subtest 86
Affixes Subtest. 91
Procedure 94
Consonants Subtest 97
Vowels Subtest 98
Inflected Endings Subtest. 98
Contractions & Possessives Subtest 99
Affixes Subtest. 99
Reading Subtest of the Metropolitan Achievement Tests. . . 100
Results 101
Phonics Test 101
Structure Subtests 114

Page

The Establishment of Performance Guidelines for the Word Identification Test Battery. 127

Conclusions. 137

Reference Notes. 141

References 143

Bibliography 159

List of Tables

Table		Page
1	Subject Population by School and Grade	72
2	Target Correspondences by Frequency and Position for Consonants	75
3	Target Correspondences by Frequency and Position for Short, Long, and Other Single-letter Vowels	81
4	Target Correspondences by Frequency and Position for Vowel Clusters	82
5	Target Inflected Endings Assessed in the Inflected Endings Subtest	85
6	Frequency of Contractions[a] and Rank Order of Contraction Categories	89
7	Target Forms in Contractions & Possessives Subtest	90
8	Target Affixes Assessed in the Affixes Subtest	93
9	Subtests by Grade Level from the Word Identification Test Battery and the Metropolitan Achievement Tests	96
10	Summary Statistics for the Consonant and Vowels Subtests	102
11	Summary of t-tests for Differences Due to Sex	104
12	Summary of t-tests for Differences Due to Grade Level	105
13	Summary Statistics for Consonants Subtest by Item Category and Grade Level	106
14	Summary Statistics for Vowels Subtest by Item Category and Grade Level	107
15	Percent of Times Each Consonant Foil was Selected	110
16	Percent of Times Each Vowel Foil was Selected	111
17	Number of Times Consonant Foils Were Selected by Each Score Group	113
18	Number of Times Vowel Foils Were Selected by Each Score Group	115

Table		Page
19	Summary Statistics for Structure Subtests.	116
20	Summary of t-tests for Differences Due to Sex on Structure Subtests.	117
21	Summary of t-tests for Differences Due to Grade Level on Structure Subtests.	118
22	Rank-ordered Listing of Mean Percent Correct on Item Categories in Inflected Endings Subtest.	120
23	Rank-ordered Listing of Mean Percent Correct on Item Categories in Contractions & Possessives Subtest. . .	121
24	Rank-ordered Listing of Mean Percent Correct on Item Categories in Affixes Subtest	122
25	Performance on Prefixes and Suffixes by Grade Level. . . .	125
26	Performance Standards for Low, Average, and High Comprehenders on Subtests in the Word Identification Test Battery	129
27	Performance Standards for Low, Average, and High Comprehenders on Phonics Subskills	130
28	Performance Standards for Low, Average, and High Comprehenders on Structure Subskills	131
29	Pearson Correlations of Subtest Category Score with Metropolitan Comprehension Scores	133

List of Figures

Figure		Page
1	Directions and practice items from the first page of the Consonants Subtest for Initial Position.	79
2	Directions and practice items from the first page of the Vowels Subtest for Medial Position.	84
3	Directions and practice items from the first page of the Inflected Endings Subtest	87
4	Directions and practice items from the first page of the Contractions & Possessives Subtest.	92
5	Directions and practice items from the first page of the Affixes Subtest	95

Abstract

The work reported in this paper culminates four years of research by the Project on the Assessment and Analysis of Word Identification Skills in Reading. The goals of the research have been to: (a) explore the relationships between the mastery of word identification skills and comprehension abilities; (b) to develop a set of diagnostic subtests which assess the word identification skills of elementary school children; and (c) to establish empirically-based mastery levels for each subtest, based on performance scores stratified by grade level and comprehension ability.

In spring 1980, the final version of the Word Identification Test battery and the Reading Subtest of the Metropolitan Achievement Tests were administered to approximately 100 children at each grade level, one through five. The data were used to examine correlations between word identification skills, as measured by the various word identification subtests, and global comprehension ability, as measured by the standardized Metropolitan reading subtest. In addition, levels of skills mastery for each of the five subskills assessed in the battery were established. This report presents some historical perspectives on word identification skills, documents the development of the test items, and summarizes the results of the analyses.

The Word Identification Test battery is comprised of five subtests within two major skills areas: phonics and structural analysis. The subtests in the battery are unique in that all target items (letter-sound

correspondences, inflected endings, affixes, and contractions & possessives) were developed, whenever possible, in accordance with word frequency information and, hence, reflect the most frequently occurring phonic and structural features of the English language.

As part of the process to establish performance guidelines for the Word Identification Test battery, an extensive review of mastery learning theory was conducted and the issue of mastery learning theory and its application to reading instruction examined. Traditionally, cutoff scores or mastery levels have been arbitrarily set by publishers of tests and appear to be absolute. For example, in many skills management programs, a score of 80 percent or better indicates mastery of a particular skill. To date, however, there has been no empirical verification that a single percentage correct score should indicate mastery for all skills. The Word Identification Test battery uses a unique approach for the establishment of mastery levels--instead of a single absolute criterion for mastery, the performance guidelines for each subtest in the battery take into account a child's grade level and comprehension ability. Using subpopulations stratified by global comprehension ability, performance standards are provided for each subtest in the battery for every grade level tested. These empirically-derived performance guidelines range from 34.3% to 96.4%, depending on the subskill being measured and the grade level and comprehension ability of the student. More eloquently than any argument appearing in the literature, this range of expected performance demonstrates the inappropriateness of arbitrary, rigid mastery scores.

INTRODUCTION

The work reported in this paper culminates four years of research by the Project on the Assessment and Analysis of Word Identification Skills in Reading. The focus of the research has been to (a) explore the relationships between the mastery of word identification skills and comprehension abilities; (b) to develop a set of diagnostic subtests which assess the word identification skills of elementary school children; and (c) to establish empirically based mastery levels for each subtest, based on the performance of groups of children stratified by grade level and comprehension ability. The test battery is comprised of five subtests within two major skill areas: phonics and structural analysis.

The subtests in the Word Identification Test battery are unique in that decisions regarding the specific information to be assessed (letter-sound correspondences, inflected endings, affixes, and contractions & possessives) were based, whenever possible, on frequency data. Hence, the subtests assess the most frequently occurring phonic and structural features of the English language. In addition, the formats of the subtests eliminate children's prior knowledge of vocabulary as a confounding factor in performance.

The mastery levels (performance guidelines) for each subtest, determined by comprehension performance and grade level rather than by arbitrary cutoff scores, will be of value to teachers for obtaining diagnostic information. The Word Identification Test battery will thus provide teachers with important information upon which to base instruction in the word identification skills most related to reading comprehension.

During the last decade, increased attention has been given to the individualization of instruction and to teacher accountability for pupils to achieve minimal competency in reading. The result has been a growing emphasis on skills development in reading. In an effort to individualize instruction, particularly in the basic skills area of reading, diagnostic testing has become more prevalent and, consequently, more time-consuming. Within the past ten years, over a dozen programs have been developed which are essentially skills management systems (e.g., the Wisconsin Design for Reading Skill Development). Furthermore, most basal reading series published during this period have included a substantial skills management component. The numerous skills management systems and basal reading programs have all made heavy use of criterion-referenced testing for assessment of the various subskills of reading.

In line with this current emphasis on the diagnostic assessment of reading skills, the present study sought to identify those word identification skills which correlate most highly with reading comprehension, to examine methods of assessing word identification skills, and to develop a set of valid and reliable diagnostic tests to assess these skills.

Over the last few years, considerable attention has been given to defining, assessing, analyzing, and teaching the three fundamental components of word identification: phonic analysis, structural analysis, and contextual analysis[1] (Johnson & Pearson, 1978). Because of the

[1] Phonic analysis: processes which help children pronounce unfamiliar printed words as an aid to understanding their meanings. Structural analysis (morphemic analysis): processes which help children determine the meanings of unfamiliar printed words by discerning their meaningful parts. Contextual analysis: processes which help children understand the meanings of printed words or phrases which are unfamiliar to them, or which help children learn new meanings for familiar words and phrases, by attending to the context of the material surrounding the given word or phrase.

current information concerning word identification has been based on speculation only, an empirical investigation of these issues is clearly warranted.

First, while reading educators agree that word identification skills are important for reading, it is not clear which correspondences, patterns, or strategies within each of the broad areas of phonics, structure, and context relate most closely to comprehension. For example, how necessary is it for children to know the 61 vowel clusters in the English language and the 2 to 14 pronunciations for each (e.g., ou as in soup, would, ground, sought)? Is it worthwhile to spend instructional time teaching the rule which governs the pronunciation of x according to its position within a word (i.e., xylophone, exam, tax)? Can we justify the numerous hours spent drawing short lines between syllables, because "better syllabicators are better readers"?

In the schools, a myriad of rules governing letter-sound correspondences and syllabication are taught, but many of these rules have exceptions or are completely erroneous. For example, the rule, "when two vowels are together, the first is long and the second silent," applies to only 45% of words at the primary level (Clymer, 1963), and to only 18% of words beyond the primary grades (Emans, 1967). This rule certainly does not hold for such words as ocean, great, bread, and pause, aunt, kraut. Of course, many rules are upheld consistently (e.g., c is pronounced as /s/ before e, i, and y, and as /k/ before most other letters)--but which of the rules are worth learning and warrant the instructional time required to teach and practice them?

Second, the issue of mastery learning theory and its application to reading instruction must be examined. Traditionally, teachers interpret scores of, say, 80% to mean "mastery" of a particular skill. This notion of an arbitrary standard is reinforced by the fact that most of the criterion-referenced tests used across the country have established cutoff scores of 70%, 80%, or 90% as indicators of mastery. To date, however, there has been no empirical verification that any single percentage correct score should indicate mastery. Moreover, there is disagreement among educators about what mastery really is. Some educators view mastery as an absolute state of proficiency; partial mastery is as illogical a concept as partial pregnancy or partial death. But it is unnatural to view the mastery of reading subskills in an absolute sense, because factors such as measurement error and attention to task must be considered on a continuum. Considering the most commonly accepted criterion-mastery level of 80%, it is justifiable to ask, "Why 80%?" And if 80%, "80% of how many of what?" And "what does mastery of a skill contribute to overall reading comprehension?"

Despite the lack of an empirically based cutoff score, the notion of mastery has strong implications for reading instruction in the classroom. An extensive review in the area of mastery learning theory was therefore undertaken by the Project (see the third section of this paper). Because the ultimate goal of reading instruction is successful reading comprehension, the mastery levels (performance guidelines) in this study were established using subsamples based on performance on a global measure of reading comprehension.

In order to individualize reading instruction, teachers must be able to effectively assess the major word identification skills. Successful individualization depends on valid and reliable diagnostic tests that delineate areas of strength and weakness for specific word identification subskills. A primary goal of the Project was the development of a Word Identification Test battery that assesses the phonics and structural analysis skills of elementary school children.

In spring 1977, a prototype of the Word Identification Test battery was developed which assessed skills in three broad areas of word identification: phonics, structure, and context. In order to provide baseline information, the battery also included a section assessing reading readiness skills. The prototype was pilot tested on a total of 282 pupils in grades one, three, and five. Following the data analysis, revisions were made on individual subtests; in winter 1977 and spring 1978, the revised Word Identification Test battery (without the reading readiness component) and the Reading Subtest of the Metropolitan Achievement Tests (Farr, Prescott, Balow, & Hogan, 1978) were administered to approximately 1,150 second, fourth, and sixth grade public elementary school children from five regions of the United States (Johnson, Pittelman, Schwenker, Shriberg, & Morgan-Janty, 1978). Following analysis of the data from these test administrations, the criteria which guided test construction were evaluated, and additional criteria were incorporated in the development of the present tests (Johnson, Pittelman, Schwenker, & Shriberg, 1979; Johnson, Shriberg, Pittelman, & Schwenker, 1979). In addition, the Project decided to limit the development of the Word Identification Test battery to the areas of phonics (with Consonants and

Vowels Subtests) and structural analysis (with Inflected Endings and Affixes Subtests). In an effort to gain as much information as possible in these areas, the Project conducted an extensive review of instructional practices and existing test instruments on phonics and structure (see second section of this paper) (Johnson, Pittelman, Schwenker, & Shriberg, 1979; Johnson, Shriberg, Pittelman, & Schwenker, 1979).

Between winter 1978/79 and winter 1979/80, the revised phonics and structure subtests were administered to several hundred pupils in grades two through five. The primary purposes of these studies were to obtain item analysis information prior to preparation of the final version of the tests and to evaluate the test directions and administrator's manual for each subtest (Johnson, Pittelman, Schwenker, & Shriberg, 1980).

In spring 1980, the final version of the Word Identification Test battery was administered to approximately 600 first through fifth grade elementary school students. The performance data were used to examine correlations between reading subskills, as measured by the various subtests in the Word Identification Test battery, and global comprehension ability, as measured by a standardized test of reading comprehension. In addition, empirically based levels of skills mastery were established for each of the reading subskills assessed in the battery.

The present report presents a review of educational practices and widely used assessment instruments in phonics and structure, a historical discussion of the issue of mastery learning theory, and documentation of the results of this final investigation.

A REVIEW OF THE INSTRUCTIONAL TRENDS AND ASSESSMENT INSTRUMENTS IN THE AREAS OF PHONICS AND STRUCTURE

As part of the procedure to develop a test battery to assess word identification skills in the areas of phonics and structure analysis, a survey of existing assessment instruments was conducted. Because the survey revealed that there were no valid and reliable instruments currently available, a review of the instructional trends for phonics and structural analysis was undertaken. It is interesting to note that while there has been a great deal of research on instruction in phonics, there appears to be a lack of agreement as to what should be taught and assessed in the area of structural analysis. A primary source used for describing the instructional trends in phonics was Word Identification--Instructional Practices: The State of the Art, by Johnson and Baumann (in press).

Phonics

Prior to 1800, reading instruction in America emphasized a strong synthetic phonics approach. Later, in the early 1800's, Horace Mann introduced the "whole-word" method of teaching reading. This new method prevailed until the second half of the nineteenth century when phonics again became popular. Rigorous phonics programs dominated the reading and language curricula from about 1880 to 1915.

Between 1915 and 1940, research on the teaching of word identification skills centered on the relative merits of a phonics versus a whole-word or look-say approach to reading. The majority of researchers in this

period who compared phonics and look-say methodologies noted superior results for instruction in phonics (Currier, 1923; Currier & Duguid, 1916; Garrison & Heard, 1931; Tate, 1937; Valentine, 1913). A classic study by Agnew (1939) found that primary grade children who received reading instruction with a heavy emphasis on phonics scored higher on tests of phonics ability, word pronunciation, oral reading, and vocabulary than did children of the same age instructed in the look-say approach.

While research tended to support the efficacy of phonics as the most efficient means of teaching word identification skills, no dominant set of instructional practices emerged. The purpose of phonics instruction is to teach children how to pronounce "unknown" words. In order for phonics analysis to be effective, however, the "unknown" word must be in a child's speaking or listening vocabulary. The assumption is that the ability to pronounce the unknown word will automatically cue its meaning in semantic memory (Johnson & Baumann, in press).

The most popular approach to teaching phonics is based on the premise that if children are able to analyze words by segmenting them into parts, they should be able to recombine (blend) these parts into new units, thereby enabling them to transfer and apply this skill in decoding unfamiliar words. Thus, the skill of segmentation appears to be prerequisite for the ability to successfully blend. Children who could segment syllables were successful in blending training, which in turn facilitated the learning of words. Research has shown that both segmentation and blending must be mastered if a phonics approach is to be successful for generalizing to the reading of unfamiliar words (Fox &

Routh, 1976; Jeffrey & Samuels, 1967; Jenkins, Bausell, & Jenkins, 1972; Muller, 1973).

The act of decoding, then, appears to be a three-stage process: children are initially taught letter-sound correspondences by analyzing words in their speaking and listening vocabularies; they are then taught to segment words into phonemic units; and finally, they are instructed in the skill of blending these isolated sounds into known and previously unknown words. It is this last step, blending, that has been shown to be the most crucial in the transfer of phonics analysis skills to the reading of unfamiliar words (Johnson & Baumann, in press).

According to Venezky and Massaro (1976), this ability to decode provides a certain degree of independence and self-assurance for beginning readers; that is, children acquire a manageable set of letter-sound associations upon which they can build a large number of words. In addition, phonics instruction, because of its emphasis on regular letter-sound associations, draws attention to the orthographically regular features of printed English words--the procedure for analyzing printed words into subunits for pronunciation facilitates acquisition of the patterns in our language which are also orthographically regular. And, in turn, because there are a limited number of ways that sequences of letters and letter groups can be put together to form English words, knowledge of this regularity can help the reader resolve the letters in a string that conforms to the language (Massaro, 1975).

With the widespread recognition of the importance of phonics in the reading curriculum, a dependable measure of phonics ability on which to base instruction is needed. The Project on the Assessment and Analysis

of Word Identification Skills in Reading has identified several issues for consideration in the development and evaluation of a phonics instrument. First, the scope of the test has to be addressed; that is, a decision has to be made as to which of the hundreds of spelling-to-sound correspondences in the English language should be selected for assessment. Next, the format of the test has to be considered: recognition response or production response, group administration or individual administration, and decoding or encoding. Finally, the modes in which the target spelling-to-sound correspondences and the response choices are presented must be determined. Pikulski and Shanahan (1980), after surveying a number of phonics tests (most of which were subtests of larger diagnostic test batteries), concluded that, to date, there was no phonics instrument available that rendered a systematic assessment of phonics skills.

A survey of the phonics components of nine popularly used tests was conducted by the Project. Following the issues identified above, several phonics tests and phonics components of diagnostic and achievement tests were evaluated: the California Achievement Tests (McGraw-Hill, 1977), the Botel Reading Inventory (Botel, 1961), the Prescriptive Readin Inventory (McGraw-Hill, 1972, 1976), the Skills Monitoring System for Reading and Word Identification (Harcourt Brace Jovanovich, Inc., 1975), the Wisconsin Design for Reading Skill Development (Otto, Miles, Kamm, & Stewart, 1972-1975), the Phonics Knowledge Survey (Durkin & Meshover, 1964), the California Phonics Survey (Brown & Cottrell, 1963), the Stanford Achievement Test (Madden, Gardner, Rudman, Karlsen, & Merwin, 1970-1974), and the Silent Reading Diagnostic Tests (Bond, Balow, & Hoyt, 1970).

ISSUES OF TEST CONSTRUCTION

Scope of Test

One important issue that was addressed by the Project concerned the selection of correspondences for assessment. Results of the survey revealed that most phonics instruments assess only a small number of the hundreds of spelling-to-sound correspondences in our language. The California Achievement Tests, for example, consist of only 25 items; 10 items assess the entire consonants category (single-letter consonants, consonant digraphs, and consonant clusters), 13 items assess single-letter vowels (all either long or short), and 2 items assess vowel clusters or diphthongs. It is questionable whether performance on only a few items should form the basis for global judgments regarding children's overall competence with phonics.

In addition to the number of correspondences to select for assessment, attention must also be focused on how often these correspondences occur in our language. Many of the tests reviewed assessed correspondences that have low frequencies of occurrence in the English language. For example, is it important to assess the vowel cluster oa as /ɔ/ (as in broad) when oa as /ɔ/ appears only 9 times in the 20,000 most common English words? Children who learn the correspondence will have little occasion to apply it in decoding unknown words. Johnson and Baumann (in press) maintain that a diagnostic instrument should reflect the information learned in the classroom. By selecting for assessment only those correspondences which appear frequently in curriculum materials, this notion of ecological validity is upheld.

There is yet another issue--word position of the target correspondence. In many of the tests surveyed, a large number of the correspondences are not assessed in the position(s) in which they most typically occur in English words. The Wisconsin Design, for example, assesses the single letter v in final word position. Data from the Venezky (Note 1) tabulations of spelling-to-sound correspondences indicate that v appears only twice in final position (although it appears 353 times in initial position) in the 20,000 most common English words. One has to consider, therefore, whether it is educationally prudent to assess v as /v/ in final position.

Format of Test

The most accurate procedure to use in assessing phonics skills is an oral productive task. The ideal phonics test would require the child to read aloud, while the examiner would record all pronunciation errors made on unfamiliar words. Pikulski and Shanahan (1980) agree that the functional use of phonics occurs when the examinee is presented with letters or words and is required to produce some audible response, a procedure generally used by individually administered tests. While an individually administered productive task would best reflect the ability to apply phonics knowledge, consideration must also be given to efficiency of assessment. Because there is no feasible way to obtain oral responses from examinees in a group testing situation, a group administered test must have a recognition rather than a production format.

The Phonics Knowledge Survey (Durkin & Meshover, 1964) is an example of an oral production test. Children view separately each of 14 consonants

and the 5 single-letter vowels and are asked to pronounce the corresponding sounds for the consonant letters, and the 5 long and 5 short corresponding sounds for the vowel letters. This productive method of assessment is not efficient, however, because of the time needed to administer the test to each individual child. The validity of the Phonics Knowledge Survey can also be questioned in terms of assessing letter-sound correspondences in isolation, rather than within the context of words.

The Botel Reading Inventory (1961), also a productive test, utilizes a written format which can be administered to a large group or class. The written format, however, puts emphasis on the encoding (sound-to-spelling), rather than on the decoding (letter-to-sound), process. Hence, spelling performance, instead of phonics ability, is being measured.

In summary, group administered tests using a recognition format are more efficient than individually administered tests using a production format. Although there is some minimal evidence to suggest that recognition phonics tasks may be easier to perform than production tasks (e.g., Guthrie, 1973; Pikulski and Shanahan, 1980) point out that there are a variety of recognition and production formats that appear to vary considerably in difficulty. In other words, there may be more variation in difficulty between different recognition test formats, or between different production test formats, than between a recognition test format and a production test format.

Presentation of Target Spelling-to-Sound Correspondences and Response Choices

Another issue central to the evaluation of an effective phonics instrument is whether the letter-sound correspondences being assessed are presented in isolation, in real words, or in synthetic words. The value of assessing sounds in isolation is questionable because (a) it is not possible to produce the sounds associated with consonants or consonant clusters without adding a vowel sound (Groff, 1977); (b) producing sounds in isolation is an incomplete activity, and is therefore not sufficiently predictive of functional ability in phonics (Pikulski & Shanahan, 1980); and (c) the sounds of many letters, especially vowels, are determined by their orthographic environments (Chomsky & Hall, 1968; Venezky, 1967).

One alternative to assessing letters in isolation is to present the target letters within a word. But the problem inherent in using actual words is that children may recognize the words as sight words and, hence, might not need to utilize a decoding strategy. The Skills Monitoring System for Reading and Word Identification (1975), for example, assesses one letter-sound correspondence for ch in the real target word, each. The four response choices are dish, Christmas, anchor, and chair. Because all four response choices include actual pronunciations for ch, children must rely on prior knowledge of the pronunciations of the target word and the response choices to arrive at the correct answer. This implies that children must recognize these words as sight words, and the correct answer is reached through auditory matching rather than decoding. Another problem in the use of real words is that the words are taken from word

lists which are often generated through random selection of words from reading materials and, therefore, may not allow for a careful evaluation of a systematic full range of phonics skills (Pikulski & Shanahan, 1980).

A second alternative to assessing letters in isolation is to present the target letters for a correspondence within a synthetic word. This format allows letter-sound correspondences to be presented within appropriate orthographic environments, and requires that children use phonics (decoding) skills rather than a sight word approach. One concern with the use of synthetic target words, however, is ensuring that the synthetic words conform to phonological principles of the English language. Sipay (1974) states that "if nonsense syllables are difficult to pronounce, or if the letter sequence confuses the learner, the examiner may be misled into concluding that the learner has weak word analysis skills" (p. 5). In the tests surveyed by the Project, the synthetic words did not always show phonological conformity. For example, in the Phonics Knowledge Survey, children are asked to pronounce the sound made by the a in the synthetic target word, aef. The correct answer is given as long a, because children are expected to apply the "rule" that when two vowels are together, the first vowel says its name and the second vowel is silent. But Venezky's (1970) tabulations show that ae in initial position is never pronounced as long a. Pikulski and Shanahan (1980), object to the use of nonsense words because "the examinee is deprived of the opportunity to match the arrived at pronunciation for a test word with a word that is a part of his or her vocabulary."

Careful attention must also be given to the development of the response choices in a phonics test. One question to address is how many

response choices should be developed for each item, because the number of response choices can affect the reliability of the test. Many of the tests which the Project surveyed had true-false, same-different, or yes-no formats, which greatly increase the likelihood that students will arrive at correct answers by guessing. Most of the tests, however, had a multiple choice format, with the number of response choices varying from three to five.

Another consideration affecting the development of response choices is the number of syllables in the words used as response choices. Many of the tests reviewed were not consistent in controlling for the number of syllables in the response choices within an item. The Wisconsin Design, the Skills Monitoring System for Reading and Word Identification, and the Stanford Achievement Test, for example, all include response choices with varying numbers of syllables. In this regard, educators (e.g. Massaro, Note 2) have expressed concern that decoding a multi-syllable word may require more complex processing than decoding a one-syllable word, and that the inclusion of both kinds of words within a given test item may confuse young children.

A third issue in evaluating response choices relates to the position of the target letters of the correspondence within a response choice word (initial, medial, or final). According to Venezky (Note 1), the position(s) of greatest occurrence varies for different letters: It seems logical, therefore, to assess a letter-sound correspondence in the word position in which it most frequently occurs. The position of the target letters within a target word should also match the position of the letters within each of the response choices. Several of the reviewed

tests include items that contain shifts in the word position of target correspondences. The Prescriptive Reading Inventory, for example, has an item in which short a is presented in medial position of the CVC word, cat. One of the response choices is the CV word, day, which has its vowel sound in final position. Young children may be required to use different, more complicated psychological processes when the correspondences of interest shift in position than when the target correspondences are all in the same position within words.

A fourth issue related to the development of response choices is the mode in which these response choices appear. All but two of the reviewed tests (the Phonics Knowledge Survey and the Botel Reading Inventory are productive tests) include response choices in the form of isolated single letters or letter clusters, real words, or pictures.

The validity of utilizing single-letter or letter cluster response choices (as in the Prescriptive Reading Inventory and the Silent Reading Diagnostic Tests) is questionable. In a review of the Silent Reading Diagnostic Tests, for example, Kress (1972) refers to the response choices as "artificial graphic representations," and is dubious about whether they truly measure the phonics abilities being assessed: ". . . in Test 6, for the beginning sound in natural, the child is to select pn, from pn, un, tn, nt; in Test 7, for the ending sound in decay, he is to select quet from khak, kayn, cove, quet; in Test 8, for the vowel sound at the beginning of the word win, he is to select i from x, i, v, a."

In the tests reviewed, the most common form of response choice is the real word (the Prescriptive Reading Inventory, the Skills Monitoring

System for Reading and Word Identification, the California Phonics Survey, the California Achievement Tests, and the Stanford Achievement Test). One of the problems with using real word response choices is the likelihood of testing visual matching, rather than decoding. In the Stanford Achievement Test, for example, children are presented with the target letter t in the real word ten. The three response choices are gate, nine, and been. Children can easily select the correct answer by visually matching the t in ten with the t in gate.

Of all the tests inspected, the Wisconsin Design was most consistent in using response choices in picture form. The use of pictures as response choices eliminates the problems associated with real or synthetic words--namely, the visual matching of real words, the recognition of real words as sight words, and the concern that synthetic words may not conform to phonological rules of the English language. In the Wisconsin Design, the examiner pronounces the picture names of the target word and of all the response choices. The focus of this test, however, is on the auditory matching of sounds, rather than on the decoding of letter-sound correspondences.

THE PHONICS COMPONENT OF THE WORD IDENTIFICATION TEST BATTERY

For the past three and a half years, the Project on the Assessment and Analysis of Word Identification Skills in Reading has been developing a phonics assessment instrument that addresses all of the issues discussed above. The final version of the Phonics Test presents target correspondences in synthetic words that are phonologically accurate and four response choices in picture form. A format consisting of synthetic

target words and pictorial response choices ensures that the instrument comes as close as a recognition test can in assessing true phonics skills, rather than visual matching of letters and auditory matching of sounds.

The Phonics Test assesses 73 different spelling-to-sound correspondences with a total of 146 items. The Phonics Test is comprised of two subtests: a Consonants Subtest, composed of 42 single-letter consonant items, 38 consonant cluster items, and 10 consonant digraph items; and a Vowels Subtest, composed of 10 short vowel items, 10 long vowel items, 8 other single-letter vowels, and 28 vowel cluster items. Each spelling-to-sound correspondence is tested with two items. Selection of target items was based on frequency data from the Venezky (Note 1) tabulations of spelling-to-sound correspondences of the 20,000 most common English words. Response choices are based on speech production data and perceptual information from the Bouma (1971), Miller and Nicely (1955), and Peterson and Barney (1952) confusion matrix studies.

Indeed, phonics instruction has been and will remain an integral part of most beginning reading programs. An effective and efficient instrument for assessing the phonics skills of primary school children was therefore developed to help teachers plan and evaluate instruction.

Structural Analysis

Structural Analysis is a strategy of word identification by which a reader determines the meaning of an unfamiliar word by identifying its meaningful parts (Robinson, Monroe, Artley, Huck, & Jenkins, 1965; Schubert, 1969; Johnson & Pearson, 1978). This process involves analyzing words and dismantling them into units of meaning (i.e., roots,

inflected endings, syllables, prefixes, and suffixes), identifying the individual meanings, and then recombining these parts into a meaningful whole (Johnson & Baumann, in press). Johnson and Bauman state that structural analysis also aids in the pronunciation of unknown words; Moyle (1974), concurs, and defines structural analysis as "a method of analyzing a printed word to determine its meaning by identifying meaningful parts . . . which in turn many be blended into the sound of the word." Although educators differ in regard to definition, it is generally agreed that the primary purpose of structural analysis is to assist the reader in looking for the familiar, meaningful parts in words that are unfamiliar as a total unit.

Instruction in structural analysis skills is aimed at helping the reader identify the meaningful units of an unknown word. By analyzing the structure of a word in this way, a reader can often approximate the meaning of a new word.

Most reading professionals recommend direct instruction in structural analysis skills (Pearson & Johnson, 1978; Farr & Roser, 1979; Karlin, 1971; Smith & Johnson, 1976; Spache, 1963; Stauffer, 1969), although there is a lack of agreement about the actual content that should be taught. Reading methods texts differ as to which skills to emphasize--some advocate an analytical approach using word configuration and context, while others promote a synthetic approach stressing letter-sound relationships and structural analysis (Witty, Freeland, & Grotberg, 1966).

According to Spache (1963), instruction in structural analysis skills should proceed from basic shape or configuration, to phonic clues,

compound words, and syllabication in the primary grades; then to roots, prefixes and suffixes in the intermediate grades. Witty, Freeland, and Grotberg view the structural analysis hierarchy as simple suffixes, compound words, prefixes, root words with inflected endings, and syllabication, dispersed from the primer to the third grade reader. In a more recent study, Otto and Chester (1976) advocate mastery of one skill level before proceeding to a more difficult skill. They identify six levels of difficulty within structural analysis skills: base words with prefixes and suffixes, singulars and plurals, syllabication, accent, unaccented schwa, and possessive forms. Educators who support the synthetic approach for instruction (Bond & Wagner, 1969; Lamb & Arnold, 1976; Johnson & Pearson, 1978) generally agree that structural analysis should encompass derivatives, variants, and compound words. Because this approach is gaining widespread popularity in language arts and reading programs throughout the United States, a discussion of these terms is presented below.

DERIVATIVES

As defined by Schubert (1969), derivatives are root words with a prefix and/or suffix. Deighton (1959) classified variant and invariant affixes into groups containing 68 commonly used prefixes and 100 commonly used suffixes, and concluded that at least two-thirds of these derivatives provide clues to word meanings. Similarly, Kean and Personke (1976) and Breen (1960) have compiled lists of affixes that should be taught, and Stauffer (1942) and Osburn (1925) each identified prefixes which warrant instructional attention.

Because derivatives affect meaning, pronunciation, and spelling patterns, there is concern as to which methods are best for teaching these forms. Should a list of variant and invariant derivatives be memorized? Should they be taught as visual units? Or, can the derivatives be learned through a combined approach?

It is generally agreed that the task of locating small words within larger words is misleading to students (Ekwall, 1970; Johnson & Pearson, 1978; Smith & Johnson, 1976; Spache, 1976) because pronunciation and meaning are affected by the letter arrangements. (Notice the "little words" in father, some, wine, potato, and honey.) But, although educators may concur about what not to teach, they are often in disagreement about what methods and skills should be taught. For example, Aaronson (1971) found that students profit most from a structured word list dictionary approach, but Spache (1976) proposed teaching derivatives only as visual and pronounceable units.

Otto and Chester (1976) support an approach which is structured and hierarchical in nature. They do not advocate memorization; instead they emphasize direct deductive instruction, which is supported by many of the major basal series (American Book Co., 1968-1972; Ginn 720, 1976; Harper and Row Publishers, 1966; Scott Foresman Publishers, 1965). Most educators (Ekwall, 1970; Johnson & Pearson, 1978; Lamb & Arnold, 1976; Smith & Johnson, 1976) would agree with Otto and Chester that memorizing long lists of derivatives is a meaningless exercise. But teaching methods such as the use of drill cards for word formation and discussion (Ekwall, 1970), subdividing unfamiliar words into meaningful parts (Smith & Johnson, 1976); building new words from familiar roots (Johnson & Pearson,

1978; Lamb & Arnold, 1976; Smith & Johnson, 1976), and reading to apply derivatives within a contextual framework (Johnson & Pearson, 1978; Smith & Johnson, 1976) are strongly advocated.

VARIANTS

Variants are words which contain a root and an inflected ending (Lamb & Arnold, 1976). Variant endings (often called inflectional endings) change a root word so that it conforms to its grammatical environment. They change verbs by time agreement, adjectives by comparison, nouns by number agreement, and adverbs by degree. Common examples of variant endings are: s, es, ed, ing, er, and est. According to Kean and Personke (1976), the eight inflections of the English language are: noun plurals, noun possessives, present tense third person singular verb, past tense verb, present participle verb, past participle verb, comparative adjective, and superlative adjective.

Variants are affected by contextual setting, and are most commonly taught within a hierarchical structure that is dependent upon degree of difficulty, grammatical class, and usage within context (Ekwall, 1970; Johnson & Pearson, 1978; Lamb & Arnold, 1976; Otto & Chester, 1976; Smith & Johnson, 1976). Although educators have differences of opinion regarding instruction in inflected endings, there is general agreement that knowledge of variants can help children analyze unknown words.

COMPOUND WORDS

A compound word is one in which two morphemes, each of which could stand alone as a root word, are combined to form one new word (Lamb &

Arnold, 1976). Helping children to identify the two word units of a compound word may serve as an early introduction to structural analysis skills (Smith, 1963).

Most basal series include lessons which introduce and give practice in identifying compound words. Johnson and Pearson (1978), however, have suggested a unique approach to the study of compound words. They recommend that children be made aware of the underlying structural relationships of compound word units, and provide a structural breakdown of six different compound word relationships:

1) B is of A: A fishbone is a bone of a fish.

2) B is from A: Hayfever is fever from hay.

3) B is for A: A dog biscuit is a biscuit for a dog.

4) B is like A: A boxcar is a car like a box.

5) B is A: A nobleman is a man who is noble.

6) B does A: A crybaby is a baby that does cry.

Syllabication is another word identification skill in which children often receive a great deal of instruction. While some educators advocate instruction in syllabication (Gates, 1947; Gray, 1960; Karlin, 1971; Osburn, 1954; Smith, 1963), the value of instruction in syllabication has been questioned by others (Deighton, 1959; Durrell, 1956; Glass, 1965; Groff, 1971; Johnson & Pearson, 1978; Spache & Baggett, 1966; Zuck, 1974). Many students appear to use the sounds represented by the word parts to determine the number of syllables, rather than vice versa. If this is the case, the use of syllabication as a word analysis tool is of little value (Lamb & Arnold, 1976). Another criticism of syllabication as an aid in word recognition is that the dividing point

between syllables is not always clear (Wardhaugh, 1966). For example, children are often taught that when dividing words into syllables they should divide vccv patterns between the consonants. But if the rule were applied to words like summer, mother, or father, the pronunciations rendered would be inaccurate. For example, in syllabicating the word father, the resultant syllables would be fat-her, leading to incorrect pronunciation of the word.

Despite differences in their approaches to teaching structural analysis, writers of methods texts, basal series publishers, and reading theorists do agree on one point: Structural analysis is an integral part of reading instruction, and the end result of such instruction should be the understanding of meaning from context. Structural analysis skills should be used (and taught) in conjunction with other word identification skills. The goal is to integrate structural analysis skills as one strategy which allows a reader to segment an unknown word into meaningful parts, and then to recombine these meaningful parts to make the total word recognizable, thereby facilitating comprehension.

STRUCTURE COMPONENT OF THE WORD IDENTIFICATION TEST BATTERY

The structure component of the Word Identification Test battery assesses two areas of structural analysis which reading theorists agree are crucial to the development of word identification ability: derivatives and variants. The battery includes an Affixes Subtest and an Inflected Endings Subtest. The third subtest in the battery, Contractions & Possessives, assesses the two uses of the apostrophe. The ability to distinguish between the two uses of the apostrophe is important for

obtaining the intended meaning of connected text which, in turn, affects comprehension.

During 1969-1974, the writing mechanics of 9-13 and 17 year old students were examined by the National Assessment of Educational Progress (1975). The uses of the apostrophe in contractions and to show possession were included among the objectives considered important for students. That understanding the two uses of the apostrophe is necessary is reinforced by Lloyd and Warfel (1972), who noted that proofreaders of newspapers, advertisements, and weekly magazines often erroneously leave the apostrophe in "its" when the possessive pronoun, and not the contraction of "it is" or "it has", is intended.

Earlier versions of the Word Identification Test battery included a subtest assessing compound words; however, the subtest was eliminated from later versions of the battery. Although instruction in compound words is a valuable aid in structural analysis, assessing a child's knowledge of compound words often becomes simply a measure of vocabulary knowledge, and the understanding of the underlying structure of a compound word is not readily transferable to unfamiliar compound words. The authors, therefore, decided that the allocation of instructional time to assessment of compound words could not be justified.

The earliest version of the Word Identification Test battery also included a subtest assessing syllabication. As discussed earlier, however, the allocation of time to assessment, and perhaps even to instruction in syllabication skills, has been questioned by many educators. Deighton (1959) summarized this position when he asserted, "To insist on mastery of 'rules' of syllabication is to make syllabication an end in itself . . .

the 'correct' way to divide a typed or printed word is of importance only to stenographers and printers, never to readers."

Inflected Endings Subtest

The selection of Inflected Endings to be assessed in the Inflected Endings Subtest was based on a review of scope and sequence charts of basal reading series, on published tests for inflected endings, and on frequency information.

To determine which inflected endings to assess, four basal reading series were surveyed: Ginn 720 (1976 edition), Houghton Mifflin (1971 edition), American Book Company (1968-1972), and Heath and Company (1968 edition). All four series prescribed instruction for the following inflected endings: (i)es, (i)ed, ing, er, s, 's (s'), and (i)est.

While most of the published tests that were reviewed assessed a sampling of inflected endings, the inflected ending items were usually incorporated into subtests which assessed other structural analysis skills. The Doren Diagnostic Reading Test of Word Recognition Skills (1973), on the other hand, had separate subtests to assess inflected endings (ing [six items], ed [four items], er [two items], and r and s [one item each]) and singulars and plurals (ies [three items], es [two items], and s [two items]). Similarly, an analysis of the Wisconsin Design for Reading Skill Development revealed a separate 12-item subtest containing two items for each of six inflected endings: ed, s, ing, 's, er, and es. However, detailed documentation of the criteria governing item selection was not provided for either the Doren or Wisconsin Design subtests.

Based on information from the review of basal reading series and published tests as well as from the survey of the literature, it was decided that the Inflected Endings Subtest should include items which sampled tense markers, adjectives, and plurals. The inflected endings selected for assessment were: s (as a plural), (e)s (as a verb), ed, ing, er, est, and y.

The number of items created for each inflected ending was in proportion to how often that ending occurred in the language. Frequency information was obtained from the Ginn Lexicon Project Frequency Listing (Johnson & Baumann, 1979), which is a compilation of four word frequency lists.[2] The 734 words in the lexicon which have a total frequency count of 300 or more were examined for the inflected endings selected for the test. A similar review was performed with the American Heritage Word Frequency Book (Carroll, Davies, & Richman, 1971).[3] A detailed description of the development of the Inflected Endings Subtest is presented in the reports The Assessment of Structural Analysis Skills (Johnson, Pittelman, Schwenker & Shriberg, 1979) and Interim Report: The Refinement of the Test Battery to Assess Word Identification Skills (Johnson, Pittelman, Schwenker, & Shriberg, 1980).

---

[2]The four word frequency lists comprising the Ginn Lexicon Project Frequency Listing are: Carroll, Davies, and Richman list (1971); Kucera-Francis list (1967); Moe Picture Book Words (1973); and Moe Oral Language Words (1974). A total of 18,979 different words are included in the Frequency Listing. The frequency count for the words ranges from 20 to 164,924.

[3]The American Heritage Word Frequency Book is a word list based on an examination of published material for children in third to ninth grade, and contains 5,088,721 tokens and 86,741 words. A total of 90 schools participated in the study, and over 5 million words of running text were extracted for analysis from 1,045 different publications.

Affixes Subtest

The selection of affixes for assessment was based on frequency information gathered from scope and sequence charts of basal reading series, from published tests of affixes, and from the SWRL Lexicon (Rhode & Cronnell, 1977).

Four widely used basal series were surveyed to determine which affixes are consistently taught to elementary school children (Ginn 720, Clymer et al., 1979 edition; Macmillan, Smith & Wardhaugh, 1975 edition; Houghton Mifflin, Durr et al., 1974-1978; Scott Foresman, Aaron et al., 1976). In most series, affixes were introduced in the beginning of second grade as syllabic word parts and form class markers. They were later reintroduced as meaningful word parts and as affixes in grades four through eight. A total of 11 prefixes and 14 suffixes were common to the instructional sequences of at least three of the four basal series.

The SWRL Lexicon (1977)[4] was also examined to determine the frequency of occurrence of words containing the affixes identified through the basal series survey. Upon completion of the SWRL frequency check, affixes with frequencies of less than 10 were eliminated from further consideration. This reduced the initial pool of affixes, derived from the review of the basal series, to 8 prefixes and 12 suffixes. One of the prefixes, non, was selected for inclusion, even though it was not

---

[4]The SWRL Lexicon is a 10,000-word lexicon of the basic vocabulary of children in kindergarten through sixth grade. It is a selective compilation of eight sources which include studies of materials written for children, materials written by children, and studies of the oral language of children (Durr, 1970; Entwisle, 1966; Green, Howard, Joerger, & Marino, 1958; Jacobs, 1967; Kolson, 1960; Murphy et al., 1957; Rinsland, 1945; and Weaver, 1955).

listed in the SWRL Lexicon. This was because non is considered to be a useful prefix and is taught in all four of the basal series surveyed. A comparison of these affixes with those included in other teaching and testing materials (Broska, Hodges, Patrick, Williams, & Oseroff, 1973; Northern Valley Schools, 1976; Otto et al., 1972-1975; Shepherd, 1973) supported the selection of the proposed list of target affixes for inclusion in the Affixes Subtest.

The selection of root words to be affixed also was carefully considered. Root words to be combined with the target affixes were chosen according to two criteria: (a) the root word should combine with at least two other affixes in order to create real word foils for the test items; and (b) the root word should be familiar to at least 70% of fourth graders, as indicated in The Living Word Vocabulary (Dale & O'Rourke, 1976).[5]

A list of potential root words--all of which frequently combine with affixes--was compiled from the fourth-grade vocabulary in the four basal series surveyed. This list was further modified to obtain at least four root words to combine with each target affix. Next, potential root words were checked in The Living Word Vocabulary for their appropriateness for fourth grade. Root words meeting the 70% familiarity criterion were then combined with appropriate affixes to create a pool of response choices.

Finally, the response choices were reviewed to insure that their

---

[5] The Living Word Vocabulary lists 43,000 words and their percentage scores based on how familiar the words are to students in grades 4, 6, 8, 10, 12, 13, and 16.

vocabulary levels were as consistent as possible both within and across test items. The final pool of target affixes developed through this process consists of eight prefixes and 10 suffixes. Further documentation of the test development is presented in the reports The Assessment of Structural Analysis Skills (Johnson, Pittelman, Schwenker, & Shriberg, 1979) and Interim Report: The Refinement of the Test Battery to Assess Word Identification Skills (Johnson, Pittelman, Schwenker, & Shriberg, 1980).

Contractions & Possessives Subtest

The first stage of development of the Contractions & Possessives Subtest was based on a two part procedure: (a) the identification of those contractions that are typically taught to second, third, and fourth grade students; and (b) a review of the formats used in the instruction and assessment of contractions. The four widely used basal series selected for review were: Ginn 720 (Clymer et al., 1979 Rainbow Edition), Heath and Company (Witty, Bebell, & Freeland, 1968 edition), American Book Company (Johnson et al., 1968-1972), and Houghton Mifflin (Durr et al., 1974-1978 edition). One skills management system, the Wisconsin Design for Reading Skill Development (Otto et al., 1972-1975), was also reviewed. A survey of these materials revealed that many of the contractions are taught by the end of second grade, and that all common contractions receive instructional attention by the end of third grade.

The next stage in developing the Contractions & Possessives Subtest was to select the contractions to be assessed, and to decide upon the

number of items needed to assess each target contraction. As with the other two structure subtests, a decision was made to base the number of target items for each contraction on frequency information. First, the contractions were grouped into categories based on which member of the word pair was contracted. For example, contractions of will, such as I'll, we'll, he'll, and they'll formed one contraction category. Next, The American Heritage Word Frequency Book (Carroll et al., 1971) (see footnote 2), was used to determine the frequencies of each of the specific contracted forms within the categories. Based on frequency tabulations of contracted forms within categories, the contraction categories were then rank-ordered and a proportionate number of specific contracted forms were selected for inclusion in the Subtest.

In addition to 21 items assessing contractions, ten items were created to assess possessives, resulting in a total of 31 items on the Contractions & Possessives Subtest. A detailed account of the development of the Contractions & Possessives Subtest is presented in the Interim Report: The Refinement of the Test Battery to Assess Word Identification Skills, (Johnson, Pittelman, Schwenker, & Shriberg, 1980).

A primary goal of reading instruction is the integration of structural analysis skills as a strategy for facilitating comprehension of the total word in context. All three structure subtests, Inflected Endings, Affixes, and Contractions & Possessives, utilize a sentence context requiring students to select a response to complete the sentence. Response foils are designed to be semantically or syntactically reasonable. To the extent that an instructional program stresses reading for meaning and skill practice in context rather than in isolation, the method of

assessment used in the Structural Analysis Subtests is particularly appropriate.

Summary

All five subtests comprising the Word Identification battery were developed after a careful survey of the literature and of existing instructional and assessment materials. It was evident from this review there was no assessment instrument currently available that addressed all of the issues discussed above. As a result the Project on the Assessment and Analysis of Word Identification Skills in Reading undertook the development of such an instrument. For all subtests in the Word Identification Test battery, selection of target items was based on frequency counts, ensuring that only those elements most frequently encountered and therefore most generalizable would be assessed. Formats were carefully designed to avoid features which might confound interpretation of test results, such as visual matching (for phonics subtests) and vocabulary knowledge (for structure subtests).

The final version of the Word Identification Test battery is a valid and reliable instrument for assessing the phonics and structural analysis skills of elementary school students. The battery will provide teachers with information with which to make important instructional decisions.

INTRODUCTION TO MASTERY LEARNING

An Historical Perspective

The history of testing human abilities reaches far back into civilized times. The first recorded testing occurred in China 4,000 years ago, when civil service examinations were administered to Chinese government employees (Popham, 1980). Assessment of human abilities in the United States, however, is a relatively new practice, originating as recently as the early 1900's. During World War I, the Army Alpha and the Army Beta (for nonreaders) were developed to assess the intellectual skills of military personnel.

Over the past 50 to 60 years, measurement specialists have followed the mental testing models established during World War I. A great effort was put forth to develop tests that would reflect aptitude and achievement in almost every subject area. Because scores from these instruments could differentiate among individuals in the content areas, such testing had its greatest application in schools.

As the population in the United States increased, students entered the schools in greater numbers and remained for longer periods of time. Educators had to develop criteria for determining which students would be eligible for promotion to higher grade levels. Many schools developed and administered criterion-referenced tests to assess subject matter considered essential for students to master. A substantial percentage of students, however, were unable to reach the absolute standards set by the test objectives. As a result, educators moved toward norm-referenced

assessment testing, whereby a student's performance would be viewed relative to the performance of his or her peers. The "average" performance of students within a particular grade became the standard level of performance that classroom teachers were urged to meet. In keeping with this standard, teachers developed instructional plans aimed at the average ability level (Westbury, 1970). In most classrooms, instruction was based on a single set of materials, objectives, and procedures, judged to be appropriate for the middle group of students.

The trend toward average-based education remained relatively unchallenged for several years. Prominent psychologists (Hall, Termin, Gesell, Kuhlman) asserted that hereditary factors limited human capacity; environmental assistance, including instructional intervention, was considered ineffective in altering nature's decision. Thus, educators were given additional fuel for their middle-of-the-road teaching designs and continued to administer the established and routine sequences of instruction.

Finally, in the 1920's, Carleton Washburne's Winnetka Plan and Henry Morrison's work at the University of Chicago's Laboratory School (Block, 1971) represented attempts to break away from average-based education towards individualized instruction. Both approaches featured specific educational objectives, carefully sequenced learning units with accompanying instructional materials, diagnostic instruments to judge student progress on each unit, supplemental corrective materials for students who needed additional help, and a flexible time schedule allowing children to progress through the units at an individual rate. But despite Washburne's and Morrison's frameworks for personalized education, the

majority of American schools were not influenced by their efforts. For nearly the first half of the 20th century, educational thinking continued to reflect the average-based philosophy, and students' abilities were assessed with norm-referenced tests.

Finally in the late 1950's and early 1960's, coinciding with the surge of scientific advancements, the measurement of abilities through norm-referenced testing was challenged (Block, 1974; Bloom, 1976). Jerome Bruner (1960), a pioneer in the new movement, proclaimed that any subject matter could be taught to any child if certain instructional adjustments were made. Support for Bruner's position came from a camp of developmental psychologists who were willing to admit that human learning could be affected by training and practice as well as by maturation.

Following Bruner, Glaser (1963) pointed out the need to determine a student's level of proficiency if appropriate instruction is to occur. Glaser suggested that since student achievement in a given subject matter ranges from no proficiency to perfect performance, assessment tools should indicate a student's exact level of proficiency and lead a teacher to identify the specific skill areas in need of further work. Glaser's thesis was that teachers needed "information as to the degree of competence attained by a particular student which is independent of reference to the performance of others" (Glaser, 1963, p. 520). Thus, by the middle of the 20th century, a new philosophy of education emerged, which focused on measuring individual learning abilities through criterion-referenced testing designs. Models of school learning were developed,

based on minimal scores from these criterion-referenced tests. The new models represent a phenomenon in education often referred to as mastery learning.

Competency-based Education

Mastery learning theory is the underlying concept of competency-based education (CBE). CBE, in its true form, is a design aimed at teaching basic subject-related skills. Skills, abilities, and attitudes designated as essential to student learning are identified and sequenced, objectives for instruction are written, and teaching and testing plans for reaching and measuring attainment of the objectives are developed. Each student's performance is monitored on a regular schedule to determine whether the objectives are being met. CBE rests on the philosophy that if appropriate materials and methods of instruction are provided, students can attain basic goals set by the school. The successful attainment of these goals, in turn, will ultimately enable the student to lead a productive life (Spady, 1977; Torshen, 1977).

In CBE, learning is a two-fold procedure: (a) minimum competencies must be set for all students to attain; and (b) provisions for advancement far beyond the minimum requirements must be designed. Once goals and objectives are established alternative materials and methods are collected to provide opportunities for diversity in teaching and learning. Time adjustments are made to allow for individual rates of learning. Ideally, instruction is offered to each learner when and for as long as it is needed.

Evaluation of a student's achievement in CBE is most often obtained by using criterion-referenced tests designed to reflect each major objective in a course of study. A standard of success is specified, although it is usually arbitrarily determined, and students are expected to reach or exceed the standard. Students who fail to meet minimum standards are guided into further individual or group work.

Despite the educational promises that the CBE approach offers, implementation has been less than satisfactory. The philosophy of CBE has been interpreted in various ways. Some schools claim to be using CBE plans when, in fact, they are using traditional instructional designs along with criterion-referenced tests. In other schools, CBE is practiced with the use of individually assigned texts, average-based instruction, and end-of-unit criterion-referenced tests. The ultimate misuse of CBE, however, comes at the state level where certification standards for student performance are established. Testing programs are adapted to measure state certification requirements, yet appropriate instructional adjustments are not made. Spady (1977) warns that states have jumped aboard the CBE bandwagon without a specific definition or plan for classroom use. With the current trends in education pushing toward identifying and measuring competencies, the definition and implementation of mastery models need to be carefully examined.

Definition of Mastery Learning

CARROLL'S MODEL OF SCHOOL LEARNING

Throughout the history of education, the basic tenents underlying mastery learning have appeared from time to time. Psychologists,

teachers, tutors, and parents, searching for ways to help children learn, have been driven by the belief that learning will occur if sensitive, systematic instruction is provided. Carroll's (1963) Model of School Learning has provided a theoretical framework that defines mastery learning projects across the United States, focuses on the teacher as the manager of children's learning:

> . . . the function of the teacher is to specify what is to be learned, to motivate pupils to learn it, to provide them with instructional materials, to administer these learning materials at a rate suitable for each pupil, to monitor students' progress, to diagnose difficulties and provide proper remediation for them, to give praise and encouragement for good performance, and to give review and practice that will maintain pupil's learnings over long periods of time (Carroll, 1970, p. 71).

Carroll's (1971) model is built on the premise that all students could achieve mastery if given enough time and optimal opportunities to learn. Carroll, however, recognizes that not all students will achieve mastery of school tasks. For example, a student's unwillingness to invest adequate time on a task is a variable that obstructs the learning process. Likewise, children will vary in the degree to which they benefit from instruction, although most students benefit from good instruction. Thus, the quality of instruction is important for school learning. Quality of instruction depends on such teacher characteristics as knowledge of the learning task, appropriate sequencing of skills, and the ability to measure a child's success at reaching objectives.

In summary, Carroll believes that school learning is possible for almost every student when a competent teacher carefully controls the learning process. Controlling the learning process includes assessment of how much learning a pupil is gaining from instruction. Carroll's model rests on frequent and precise measurement. Frequent testing, administered as a functional part of instruction, should provide feedback for teachers to use in planning corrective lessons or in advancing a student to the next prescribed stage in the learning sequence. Discrete and precise instruments, designed to present items which probe the mastery of stated instructional objectives, are central to the successful implementation of Carroll's model.

BLOOM'S THEORY OF SCHOOL LEARNING

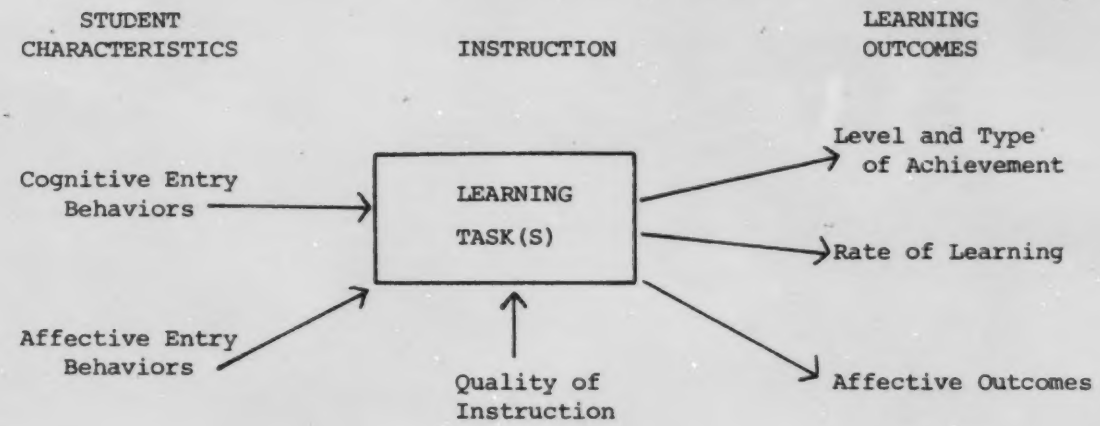
According to Bloom (1976), the goal of education should be to help all people attain the highest quality of life possible through promoting full development of each individual citizen: "what any person in the world can learn, almost all persons can learn if provided with appropriate prior and current conditions of learning" (p. 7).

Drawing heavily on Carroll's model, Bloom established a theory of school learning aimed at achieving this goal by attempting to predict and explain what he believes should happen in the process of education. Bloom (1971) interprets Carroll's position in the following way:

> . . . if students are normally distributed with respect to aptitude for some subject and all students are given exactly the same instruction (in terms of amount and quality of instruction and learning time allowed), then achievement

measured at the subject's completion will be normally distributed. Under such conditions, the correlation between aptitude and achievement will be relatively high (r = + .70 or higher). Conversely, if students are normally distributed with respect to aptitude, but the kind and quality of instruction or learning time allowed are made appropriate to the characteristics and needs of each learner, the majority of students will achieve subject mastery. The correlation between aptitude and achievement should approach zero (p. 50).

Going beyond the Carroll model, Bloom's theory incorporates student characteristics, instruction, and learning outcomes. These interdependent, alterable variables are diagrammed in the model below:



Learning Tasks and Units

A central issue underlying Bloom's theory was the development of learning tasks. In order for the theory to gain acceptance in school settings, the tasks had to be adaptable to: (a) group or individually based instruction; (b) traditional or "open" classroom settings; and (c) several types of instructional materials and teaching styles. A learning task was defined as a unit of subject matter requiring between 1 and 10 hours of a student's time. In terms of implementation, a teacher was expected to examine an entire course of subject material and divide it into small units of instruction. Sometimes it was necessary to sequence the tasks in a hierarchical fashion. Each unit, equipped with a fixed set of objectives, was presented to the student--ideally, at a rate commensurate with his or her capabilities and under optimal instructional conditions.

Strategies for feedback and corrective assignments are inherent in the unit task plan. Students are expected to successfully complete one task before moving ahead to the next level of coursework. Because students progress at different rates, Bloom (1971, 1976, 1978, 1980) has attempted to explain this variance in achievement in terms of his model.

Cognitive Entry Behaviors

Successful completion (mastery) of a learning unit depends to a great extent on what the student brings to the task. In summarizing various short-term and longitudinal studies on cognitive entry behaviors, Bloom (1976) estimates that up to one-half of the variance in achievement

can be accounted for by noting student abilities at the outset of a learning task. Specific task-related skills and general learning abilities (communication skills, learning styles, and so on) together form the total package of prerequisite behaviors. Readiness to learn also has a powerful impact on a student's success in school at all levels.

Affective Entry Behaviors

Attitude toward or interest in a learning task has an influence on achievement. A student beginning a unit of material with a positive attitude is more likely to reach a higher level of achievement than a student with a negative attitude. Based on research attempts, Bloom (1976) claims that the causal link between attitude and achievement may explain about one-fourth of the variance in achievement scores. Two types of affect appear to develop as a student progresses through the educational system: subject-related interest and attitude toward school learning in general. According to Bloom, children are not born with a set of affective characteristics; hence it is the teacher's responsibility to motivate students in the content areas. Positive feelings in specific skill areas contribute to the total impression a student has of school learning.

Quality of Instruction

In addition to what students bring to the learning task in terms of readiness and attitudes, the quality of instruction can have a significant effect on the learning of school tasks: "Who can learn in the schools is determined to a large extent by the conditions in the

school; the quality of instruction is a major determiner of who will learn well--the few or the many" (Bloom, 1976, p. 438).

Learners vary in the type and amount of instruction needed for successful performance. Bloom defines good instruction in terms of four aspects of teaching: (a) cues as to exact student responsibilities; (b) opportunities for active student participation in learning; (c) reinforcement when successful learning occurs; and (d) feedback on tasks completed and corrective assignments when necessary. High quality instruction, desirable at any point in a student's career, is especially important in the formative stages when basic skills and affective characteristics are developing.

Affective Outcomes

According to Bloom (1978), one of the most important outcomes of learning is the influence of the affective domain on the student's future achievement. An individual's perception of him- or herself as a learner in a content area not only influences school-related achievement in that area, but can also have a long-range influence by encouraging or limiting career choices. Moreover, the overall perception a person develops regarding achievement across subject areas affects his or her self-concept, and perhaps even general mental health.

Rate of Learning

A student does not progress at the same rate during all stages of learning. Adjustments in time need to be made as the student progresses through successive learning units. Early units in a sequence usually demand greater variations of time and adaptive instruction than later

units, because entry skills and attitudes are different. With good instruction and proper motivation, Bloom's theory (1976) proposes that variation in learning rates will decrease as a student approaches the final units in a sequence. Good instruction builds the requisite cognitive entry behaviors that permit students to achieve unit objectives. Achievement of objectives helps students to develop positive attitudes toward the particular unit and toward themselves as capable learners.

Level and Type of Achievement

In review, Bloom has developed a theory of school learning that has influenced the development of thousands of mastery learning projects throughout the world. School learning, according to Bloom (1980), is a result of the interaction among the student's cognitive background, the student's attitudes and interests, and the quality of instruction provided by the teacher. At each level, a student's present ability to learn is determined by previous learning along with the quality of instruction that enhanced or discouraged the learning. The anticipated outcome of the theory is that most individuals can learn--that is, achieve mastery--if given sufficient time and appropriate instruction. If mastery learning theory is effectively implemented into classroom practice, Bloom (1971) concludes that at least 95% of all students should achieve objectives established by the school. Variance in school achievement should narrow as the model is applied successfully.

Mastery Learning in the Classroom

TWO BASIC INSTRUCTIONAL DESIGNS

Mastery learning strategies have been applied at every grade level in schools throughout the world. Block (1974) has written a descriptive and comparative summary of the two most popular approaches used in applying mastery learning theory: Bloom's (1968) "Learning for Mastery (LFM) Program" and Keller's (1968) "Personalized System of Instruction (PSI)." Both Bloom and Keller have designed basic plans for implementing mastery learning theory in the classroom. Inspection of the features of each of the two designs will reveal their similarities and differences.

Bloom's Learning for Mastery Program

Mastery learning in the classroom, according to Bloom's LFM approach, involves:

1. Well planned group-based, teacher-paced lessons aimed at minimizing the amount of time needed for achieving instructional objectives.

2. Learning units, devised and sequenced by the teacher, which include instructional objectives requiring about 2 weeks of student effort.

3. Teacher-developed, diagnostic-formative tests administered frequently to assess each student's progress.

4. Corrective assignments, based on alternative materials and activities, to provide learners with opportunity to achieve objectives missed during the regular course of instruction.

5. "Mastery" achievement at each step of a unit prior to moving ahead to the next portion of study.

6. A final "summative" examination over all course objectives after all individual units in an instructional sequence have been taught. The criterion for mastery is set at between 80 and 90%. The results of this final test will determine a student's course grade.

Keller's Personalized System of Instruction

Keller's PSI approach is best described as a programmed system of mastery learning.

1. Course objectives are divided into learning units; not more than 1 week of student time is required for each unit.

2. The teacher establishes procedures for students to follow to master each unit. These procedures include study hints and guides, written materials, and a test of the unit material.

3. Students are directed to proceed through the units on a self-paced schedule.

4. An examination is given following completion of each unit.

5. A student who fails a unit examination is required to review the same materials and to retake the examination for that unit. Review and retesting continue until the unit's objectives are met with 100% accuracy.

6. Final course grades are determined by the number of units an individual has completed. A specific number of units is set as a prerequisite for a passing grade.

NOTED EFFECTS OF MASTERY LEARNING THEORY

Over a decade has passed since schools began implementing mastery learning strategies. Because school systems throughout the world are experimenting with mastery learning programs, a great deal of information exists on the effectiveness of mastery learning (Block, 1974; Block & Burns, 1977; Bloom, 1976; Kulik, Kulik, & Cohen, 1979; Torshen, 1977). Block (1979) notes, however, that research interests have shifted in the last few years from looking for evidence of the effectiveness of mastery learning to searching for an understanding of why the strategies work.

Research on the Components of Mastery Learning Models

Cognitive Achievement. Probably the most frequently cited research efforts on the relationship of student achievement to mastery learning have been those involving a large number of Korean school systems (Block, 1974; Bloom, 1976; Torshen, 1977). Two Korean educators, Kim and Lee, along with their colleagues, designed experiments to implement and evaluate mastery learning strategies. Thousands of Korean students from rural and urban areas were equally divided into mastery learning instruction and traditional instruction (control) groups and were taught in several content areas. Results of the Korean investigations consistently demonstrate the effectiveness of mastery learning research strategies in producing positive cognitive growth. For example, Torshen's (1977) summary of the Kim-Lee efforts alluded to a study involving 5,800 seventh graders from several middle schools. The students received instruction in English and mathematics under either a mastery learning or a nonmastery

(traditional) curriculum plan for 8 weeks. The minimum criterion for mastery on postassessment measures was 80% correct. The following results were obtained: In English, 72% of mastery students and 28% of nonmastery students met the criterion; in mathematics, 61% of mastery students and 39% of nonmastery students met the criterion.

Retention of Learning. Students' retention of the knowledge and skills learned in school has been the object of numerous research studies. The ability to reach a predetermined mastery score is without value if the knowledge is not retained and available for future application. Most retention studies have a four-part design: (a) a unit of material is presented to students; (b) a postassessment summative test is administered; (c) the same postassessment instrument is administered after a time lapse of a few weeks to one or more years; and (d) scores on the two identical tests are compared to note levels of retention. Two trends can be noted in the results. First, students in mastery learning programs did exhibit greater levels of retention than matched groups taught under nonmastery programs (Block, 1972; Romberg, Shepler, & King, 1970). Second, students who performed at higher levels of mastery (90% or above) tended to exhibit greater retention (Anderson, Scott, & Hutlock, 1976; Block, 1972; Poggio, 1976).

Transfer of Learning. Another area of concern is whether mastery learning approaches aid students in the transfer of learning from one class to another or from school-related to extracurricular situations. Reports of several studies summarized by Block (1972) suggest that students involved in mastery learning programs from kindergarten through college level are successful at applying previously learned knowledge

to new courses of study. The notion of transfer is especially important to the concept of mastery learning because a student must master one unit prior to moving on to the next level. If mastery learning programs did not promote transfer of learning, this unit-by-unit requirement would be unwarranted (Block, 1974).

Affective Characteristics. Affective characteristics of the learner are a major concern in the mastery learning programs based on Bloom's model. As in all areas of mastery learning research, results of studies measuring the affective domain must be viewed as tentative. Many instruments available for measuring the affective domain are limited in scope or are appropriate only when students are at certain developmental stages. In addition, the relationships between the affective domain and school achievement are complex and controversial (Torshen, 1977). Because teacher characteristics have a marked influence on children's attitudes toward learning, teacher attitudes, as well as student attitudes must be considered.

Reports by teachers involved in mastery learning projects have generally been positive (Barber, 1979; Hyman & Cohen, 1979; Torshen, 1977). Many teachers feel comfortable with the stability of the controls built into mastery programs (Anderson et al., 1976). It also appears that working with specific objectives and related materials increase teachers' confidence in their own ability to teach. Teachers in mastery learning classrooms seem to set higher expectations for their students because it is inherent in the program that students work toward a minimum level of competency.

Students enrolled in mastery learning classrooms indicate favorable attitudes toward coursework and school in general (Anderson et al., 1976; Block, 1972; Bloom, 1976), although less favorable attitudes were evident when the criterion for mastery was raised to 95% and higher (Block, 1972). Attitudes toward coursework and school ultimately relate to academic achievement; students who exhibit positive feelings are apt to spend more time in study, and, hence, become more successful learners.

Rate of Learning. Rate of learning has been defined as the time a student devotes to a learning task. In order to receive the maximum benefit from instructional activities, however, a student needs to be actively involved in the task. Mastery learning programs appear to increase the amount of time students spend actively engaged in learning tasks (Anderson et al., 1976; Hyman & Cohen, 1979) and may, in fact, help students make more efficient use of their time. Because consistent feedback is an integral component of mastery learning programs, students may experience an added incentive to complete tasks on time (Torshen, 1977).

ADDITIONAL CURRENT RESEARCH SUMMARIES

Since 1963, Hyman and Cohen (1979) have implemented and monitored Learning for Mastery (LFM) programs in reading and mathematics in over 3,000 schools. Ten pedagogical conclusions suggested by the authors are paraphrased below:

1. LFM was found to be consistently more effective in the attainment of competencies than traditional curriculum (supported

extensively by Block, 1973).

2. The effects of LFM, rather than its effectiveness, should be examined. Research is needed to investigate specific questions related to what causes LFM models to be successful in helping students reach competency requirements.

3. Increasing "time-on-task" increases the likelihood that a student will achieve mastery; in fact, the best predictor of performance is the amount of time students spend on learning tasks.

4. LFM students master more instructional objectives at a faster rate than students in non-LFM classrooms, because objectives are carefully defined and students must continually demonstrate movement toward mastery.

5. Mastery can be increased through active student participation which is aided by: (a) carefully designed behavioral objectives which guide the student and teacher; (b) direct teaching of identified objectives; (c) providing immediate feedback to the student; (d) maximizing use of positive feedback to instill a high self-concept; (e) minimizing the size of unit tasks to promote closure; (f) controlling the materials students use; and (g) positively reinforcing the learner's correct responses.

6. Individualized LFM methods are more effective than group LFM methods.

7. The popular notion of competency-based instruction (CBI) may or may not include the LFM model. Several CBI programs are merely lists of objectives with accompanying tests.

8. The goals of LFM are met only if students carry their skills out into the world and use them efficiently and effectively.

9. LFM is learning-oriented, whereas most school programs are teaching-oriented.

10. Classroom teachers can easily be trained to manage LFM classrooms.

Burns (1979) carefully examined research reports on mastery learning projects that had been collected and synthesized by Block and Burns (1977) and by Kulik et al. (1979). The authors had presented collections of research on three components of learning outcomes: cognitive achievement, retention, and affective achievement. Burns concluded that for each component, the results favored mastery strategies over traditional methods of instruction. Nevertheless, the question of whether "mastery strategies work equally well for different kinds of learning and for different types of students" still needs to be addressed (Burns, 1979). LFM designs have been accused of promoting lower-level cognitive tasks while ignoring higher-level learning. The research is unclear as to which types of learners will benefit most from a mastery learning approach.

Although most of the publicity on LFM has been favorable, some negative comments have also been noted. Glickman (1979) for example, questions a basic LFM tenet that most students have nearly the same potential to achieve that which the schools have to teach. He notes that research by Piaget, Bruner, and Elkind emphasizes the unique qualities and developmental rates exhibited by individuals. Glickman is also concerned that children who are forced to spend excessive time

mastering specific unit objectives may be denied the opportunity to benefit from more appropriate and essential developmental tasks. Finally, Glickman fears that LFM models, in aiming to develop equal skills among all students, deny a basic premise of democracy--that is, to encourage and promote the development of unique qualities within each individual.

Mastery Learning Theory: Some Applications to Reading Instruction

During the past decade, pressures from parents and the business community have forced educators to look carefully at what children are learning, how much they are learning, and the appropriateness of their school learning for real-life problems and needs. Reports of accountability, minimal competencies, and competency-based education have flooded the popular press and professional journals. Societal demands on education have influenced 34 state legislatures to mandate the development of minimal requirements to be measured by competency testing programs (Koenke, 1979; Rupley & Longnion, 1978).

Instruction in reading has been affected by the new demands placed on educators. A survey of literature on reading throughout the 1970's reveals numerous references to objective-based reading instruction and to the establishment of minimal levels of reading proficiency required for graduation. Changes have been made in instructional materials and objectives, teaching methods and techniques, and assessment procedures. The rationale underlying these changes is that each student be given the opportunity to develop to his or her full potential in reading, and

thus be able to contribute to the growth of a stronger society. Bloom's theory of mastery learning has provided the impetus for such experimentation in reading education.

MASTERY LEARNING AND READING INSTRUCTION

Freebery (1978) applied Bloom's theory to the development of a reading program in Florida. Fourteen students, scoring below grade level and classified as disciplinary problems, were presented with reading instruction based on mastery learning strategies. Freebery concluded that both improvement in reading achievement and discipline occurred in part from using the mastery learning approach.

Blohm (1978) designed a study to test the hypothesis that (a) reading comprehension improves with the teaching of subskills, and (b) testing of subskills would affect learning on a short-term basis. A group of 500 tenth-grade students were divided into two experimental groups and one control group. One experimental group received instruction on reading subskills followed by mastery testing; a second group was tested for mastery of subskills but received no special instruction; and the control group was not given any special instruction or testing in subskills. An analysis of data gathered on a delayed posttest revealed that the group receiving both instruction and testing in subskills scored significantly higher in reading comprehension than the other two groups. Although Blohm did not adhere strictly to the mastery learning approach, the results of his study indicate that mastery of subskills does have a positive effect on reading comprehension.

One of the most extensive applications of mastery learning theory to reading instruction has been the Chicago Mastery Learning Reading Program (Hannon, 1979). Chicago's mastery learning program, initiated in 1975, faced complex political, social, and financial obstacles (Katims, 1979). Among the problems were extreme cultural diversity of students, pupil:teacher ratios nearing 35:1, and a limited budget. After the program had been in effect for only 1 year, however, Smith and Wick (1976) reported five positive results of the mastery approach to learning:

1. Pupil rate of learning increased by 20%.
2. Higher achievers did not exhibit a decrease in learning rate.
3. Variance among pupil scores decreased.
4. Correlations between prior ability and performance on formative tests declined.
5. Teacher enthusiasm was high.

Furthermore, Katims, Smith, Steel, and Wick (1977) analyzed the results of the Iowa Test of Basic Skills and reported that pupils in Chicago receiving mastery learning instruction in reading had greater increases in their scores than children in the control groups.

In summary, mastery learning designs have been successfully applied to instruction in reading and could become a way of meeting public pressures to demonstrate student achievement. One component of the Chicago mastery plan was a management system used by teachers and administrators to monitor read-ng skills development. Objective-based management systems have become increasingly widespread; examination of these systems reveals that the philosophy underlying their development

is very similar to that of mastery learning theory.

OBJECTIVE-BASED READING INSTRUCTION

Objective-based reading programs are generally intended to supplement basic instruction. Usually the systems include (a) an identification of subskills essential for competence in reading, (b) a listing of objectives that must be taught and measured, (c) criterion-referenced tests designed to assess students' skill development, (d) sources of materials to use in teaching skill lessons, and (e) techniques for recording progress in skill development (Stallard, 1977b). The assumption behind these management programs is that reading is a measurable entity and mastery of individual subskills will contribute to overall reading achievement.

The use of management systems nearly parallels the mastery learning strategies discussed earlier. First, students are given tests to determine specific reading skill needs; next, prescriptive teaching and learning occurs; and finally, posttests are given to determine the efficacy of the teaching and learning. Children who have mastered a set of objectives move ahead; those who fail to achieve mastery are given additional time and corrective instruction. Mastery levels on criterion-referenced tests, usually established by the publisher, tend to center at about 80%. (See Stallard, 1977a, for an analysis of 15 widely used management programs.)

Objective-based reading instruction has been a controversial issue among reading educators. Proponents (Duffy & Sherman, 1977; Otto &

Chester, 1976; Samuels, 1976) point to the strengths of this educational innovation in terms of measuring individual skill development, focusing instructional time, and recording and reporting student progress. Opponents (e.g., Bagford, 1977) question the validity of the tests used to determine skill mastery, the feasibility of identifying a hierarchical arrangement of reading subskills, and whether reading can be segmented into a myriad of subskills.

Duffy (1978) acknowledges the shortcomings of objective-based instruction when carried to an extreme, but argues that it can be very effective when used in moderation:

> Objective-based instruction is in essence a strategy for organizing the nuts and bolts of the reading curriculum into manageable systems useful to teachers and pupils. Carried to extremes or applied inappropriately, it can lead to disaster since reading is too complex to be completely captured in a set of skill objectives and teaching is an art requiring more than mere testing and teaching of skills. However, when applied flexibly and with a sense of balance, objective-based instruction can be a useful tool in the teachers repertoire (p. 522).

MINIMAL COMPETENCIES IN READING

The era of accountability, with its demands for concrete evidence of mastery, has been instrumental in establishing and testing minimal competencies. Well over half of the states require minimal skills or "basic literacy" as a requirement for high school graduation.

The issue of minimal competency testing has been met with much emotion by reading educators. In April 1979, the International Reading Association warned that a single assessment of minimal competency should never be used to determine student promotion or graduation. In the IRA Board of Directors' recommendation, a statement was issued recommending that instead of a single instrument, decisions be made using a variety of diagnostic tools and that efforts be made to remedy deficiencies based on the diagnosis.

Seymour (1979), taking issue with the IRA publication Minimal Competency Standards: Three Points of View (Goodman, Farr, & Cassidy, 1978) proposes that schools should be held accountable for skill development, and that government mandates should pressure schools into producing more capable graduates. Seymour attributes student failure to the absence of standards and to the lack of student concern in the schools. He calls on all those concerned with education to "devise and institute minimum standards that will help raise the competence of students to levels indicating mastery of the basic skills" (p. 220).

McDonald (1978), another proponent of minimum competencies testing, expresses concern about the influence that parent groups, school boards, and state legislatures have had in establishing minimal competency standards. He cautions reading specialists to take the leadership roles in directing the development of those competencies.

Tierney (1978), an opponent of mandating minimum requirements and of the accompanying competency testing programs, expresses concern that the minimum may become the maximum requirement of students. He believes that too many unanswered questions exist regarding testing, student

retention, and the improvement of basic skills. Farr and Roser (1974) also question the rationale behind the growth of interest in testing and suggest that tests are given to find out how well the educators have done their jobs. Cassidy (1978), another opponent of establishing minimal competency standards, addresses the issues raised by Purves (1976): (a) the doubtful validity and reliability of the tests used to assess reading competencies; (b) the fear that teachers will teach to the tests; and (c) a lack of concern for cultural and language diversity. Each of these issues reflects the belief that minimal competency teaching and testing require careful examination by educators.

Demands by pressure groups throughout society have forced educators to face the accountability issue. Responses have taken the form of mastery learning strategies, objective-based reading programs, and the establishment of minimal competency standards. All three responses seem to have similar roots: Identify what needs to be learned and establish related objectives, develop a way to instruct and assess children on the objectives, and provide feedback and corrective assistance so as many students as possible reach an established standard of achievement.

The Measurement and Evaluation of Mastery Learning

The intent of this final section on mastery learning is to discuss measurement and evaluation. A basic question is, "How do we determine when a child has truly mastered a skill?" Information relevant to this question has been divided into three categories: (a) determining the

appropriate type of assessment instrument; (b) developing adequate testing devices; and (c) deriving, interpreting, and using test results.

DETERMINING THE APPROPRIATE TYPE OF ASSESSMENT INSTRUMENT

In order to make sound instructional decisions, it is necessary to determine whether a child has mastered a particular skill. Information on mastery or nonmastery provides instructional cues to the teacher and the student--that is, mastery indicates that the student should progress to the next unit of study; nonmastery indicates that additional time and corrective instruction are needed. Accurate information regarding a student's proficiency on specific subskills is generally obtained from performance data on an appropriate assessment instrument. Although both formal and informal assessment measures provide the teacher with valuable data, the present discussion will address formal means of assessment only--specifically, norm-referenced and criterion-referenced tests.

Norm-referenced Tests

According to Popham (1978), "a norm-referenced test is designed to ascertain an examinee's status in relation to the performance of a group of other examinees who have completed that test" (p. 24). In other words, a norm-referenced test permits the examination of an individual's test score in relation to the scores of his or her peers.

Norm-referenced tests (NRT) provide helpful information to educators who need survey information or comparative growth data. Items selected for inclusion in a norm-referenced test are usually of average difficulty

level. This results in a large spread of scores, or high response variance, that is essential when comparisons are to be made (Farr & Roser, 1974). Norm-referenced tests are especially useful for ranking students in terms of aptitude, predicting students' potential, and making comparisons between groups or individuals.

Definition and Uses of Criterion-referenced Tests (CRT)

Popham (1978) states that "a criterion-referenced test is used to ascertain an individual's status with respect to a well-defined behavioral domain" (p. 93). Criterion-referenced testing is based on "the notion of a continuum of knowledge acquisition ranging from no proficiency at all to perfect performance" (Glaser, 1963, p. 519). An individual's score on a criterion-referenced test is an indication of ability at a particular point in time on a specific unit of material. A score on a criterion-referenced test measures a student's ability (i.e., "How much does this child know?") in relation to a set of objectives. Results of a criterion-referenced test can be used to classify examinees as "masters" or "nonmasters" of an objective in order to plan the next step of instruction (Berk, 1980).

The Development of Criterion-referenced Tests. Several essential features must be included in the development of a criterion-referenced test. First, the test manual should contain detailed descriptions of the purported objectives of the test. In addition to helping teachers understand what the test intends to measure, this list of specific objectives is more apt to result in good item development. Because it is not possible to include items which reflect every aspect of an

objective, a writer of a CRT must choose only those types of skills, from a pool of alternatives, that best reflect mastery of an objective. Once the appropriate skills have been designated, other test specifications can be prepared: (a) general descriptions of the behaviors to be measured; (b) sample items; (c) a list of stimulus attributes (characteristics of the "stems" or stimuli used to lead children to selecting an answer); (d) a list of response attributes; and (e) specifications for any supplementary materials. The validity and usefulness of the resulting CRT will depend on how carefully the test specifications have been considered. Second, the items selected for inclusion should be representative of the entire domain of behaviors to be measured (Berk, 1980). The items must reflect their respective objectives and discriminate between groups of masters and nonmasters.

Appropriate test length is difficult to determine. If too few items are included, test results may be unreliable; on the other hand, too many items render the instrument cumbersome and inefficient. Hambleton, Swaminathan, Algina, and Coulson (1978) point out that the number of test items used to measure each objective will reflect the usefulness of the test score. If the number of items is insufficient, the decision leading to the mastery or nonmastery classification will be inconsistent in test-retest situations. Lengthening the test will lower the chance of miscalculating a student's status; however, lengthening the test means decreasing instructional time and the tradeoff may not be worthwhile.

In contrast to the norm-referenced test, in which the selection of items is designed to yield a high response variance providing a spread

of scores, criterion-referenced measures are not designed to exhibit wide variance in performance. Techniques for determining reliability, however, are available despite the limited variance. Reliability is often established by test-retest score constancy techniques. The issue of validity must also be considered in the development of criterion-referenced tests. Popham (1978) suggests three validation strategies: submitting test items with lists of test specifications to a panel of experts, checking the outcomes with test predictions of success, and systematically evaluating all possible domains which might affect mastery of the objectives.

THE ISSUE OF PERFORMANCE STANDARDS

In the mastery testing arena, it is expected that a standard for performance be established. Such a standard, often referred to as a mastery level, a cutting score, or a minimum pass level, is used by teachers to determine a child's success or failure on a unit or set of objectives. Although the establishment of performance standards has been a part of educational evaluation for over 20 years (Torshen, 1977), at the time of this writing no empirically based guidelines for establishing mastery performance standards exist (Berk, 1980; Block, 1974; Bloom, 1976; Popham, 1978; Popham & Baker, 1970; Terwilliger, 1979; Torshen, 1977).

After using a particular testing instrument with several groups of learners, experienced teachers will intuitively know what scores are necessary for mastery; some flexibility in setting acceptable performance standards is permissible. Establishment of minimal levels of competency

has depended on human judgment. As Popham (1978) noted, it is not as if there were a "true and definitive minimal proficiency level lurking out there if we [were] only clever enough to ferret it out. Minimal performance levels will always be judgmentally based, and hence, subject to the frailties of human judgment" (p. 167). Popham points out, however that performance standards must not be arbitrary or "off the wall," but, instead, should "rely on recent collateral data, wide-ranging input from concerned parties, and systematic efforts to make sense out of relevant performance and judgmental data" (p. 169).

Fortunately, progress has been made in establishing performance standards, although the issue of setting mastery scores using systematic procedures is far from resolved. Both the continuum and state models summarized and discussed by Meskauskas (1976) have generated research designed to develop quantitative models of standards setting. Continuum models are based on a belief that each learner is at some point along a path of knowledge acquisition and that a student's score is an indicator of his or her present level of learning. State models, on the other hand, describe mastery status in definite terms. In state models, there is no room for "partial mastery" because mastery is defined as complete knowledge. But, in both continuum and state models, human judgment remains a factor in establishing standards for mastery.

Hambleton et al. (1978) have discussed the issue of determining mastery states. These authors criticize the practice of comparing an examinee's "domain score" to an established cutoff score on a criterion-referenced test, thus classifying the examinee as a "master" or "nonmaster." They (1978) suggest classifying examinees as masters, partial masters,

or nonmasters. Swaminathan, Block, and Ravitch (1972) proposed that categories of cutoff scores be established and that pupils be assigned to instructional settings appropriate to their performance. Kriewall's (1972) model also has students categorized into groups along the mastery continuum in what he terms "proficiency distributions."

Huynh (1976) applied a decision-theory framework in an attempt to assign examinees to mastery status. Huynh's work also concerns Alpha and Beta measurement errors based on the use of domain scores. At present the studies of Huynh (1976) and Hambleton et al. (1978) are in a developmental stage and must be investigated at length before putting them to practical use.

Millman (1973) proposed two procedures for determining cutoff scores. First, he suggested the cutoff score be set at the point where a predetermined percentage of a given group of students would pass (or be considered masters). This procedure has been rejected because it defied one of the basic tenets of mastery testing--that an individual be evaluated in terms of his or her personal performance on a set of questions reflecting specific objectives. Millman's second suggestion was to develop a criterion-referenced test for collecting a set of scores from students who had already mastered a group of identified objectives. A raw score selected from this group would then be established as the "cutting score" for the test. But Millman also believes that mastery levels do not have to remain as absolutes. Higher cutoff scores may be feasible for fundamental skills; on the other hand, the establishment of mastery levels may not be warranted for nonessential skills. Adjusting or lowering cutoff scores is suggested when remediation costs become

prohibitive or the psychological effects of mastery learning become taxing for the individual (Terwilliger, 1970).

Block (1972) designed a study to examine the effect of varying cutoff scores during the course of instruction. Students who were expected to achieve higher cutoff scores performed better on achievement measures, transfer of learning, and retention of knowledge and skills. In addition to these academic improvements, affective behaviors showed an increase until the minimum pass level approached 85%. Results from the Block study suggest that varying cutoff scores influences performance and attitudes. Hence, it may be necessary for teachers to alter expectations for particular groups of children.

The establishment of performance standards is important to student achievement (Block, 1972, 1973). Imposing standards provides students with motivation and results in increased achievement scores, especially for those students with inconsistent study habits. Torshen (1977), however, warns against excessive enforcement of inappropriate standards. Imposing difficult or impossible standards could prohibit slow students (perpetual nonmasters) from adequate exposure to school learning. According to Torshen, when a student fails to achieve mastery even when feedback and corrective procedures have been offered, teachers should: (a) examine the objectives carefully to determine if mastery is essential to the student's future learning; (b) consider an alternative means of instruction that does not require mastery of the objective; or (c) move the student ahead to a new area of work while continuing to offer assistance in the area of difficulty.

Terwilliger (1970) raises an interesting issue regarding criterion scores in mastery learning. If a mastery level of 80% is the performance standard on a 10-item test, there are 56 ways of obtaining mastery (8 or more items correct). As the number of test items increases, the possible combinations of items that could equal mastery increases drastically. Maintaining the criterion of 80%, but increasing the number of test items to 20 would yield 6,196 possible response combinations; raising the number of test items to 30 would yield 768,212 combinations (Terwilliger, 1970). Thus, the ambiguity of test results increases with test length.

Terwilliger (1979) has suggested that teachers adopt a compromise plan that involves either an adjustment of the mastery criterion level, or a lowering of the quota of students expected to reach the set standard. By making these adjustments, Terwilliger maintains that the positive aspects of the mastery model can still be retained in classroom instruction. For example, when examining course objectives, teachers should determine which objectives all students should master and which are advanced, complex objectives that only a portion of the students can be expected to master.

In summary, it appears that mastery testing is best served by criterion-referenced measures. Those measures, however, must be carefully developed according to explicit test specifications based on the stated objectives of a learning unit. The establishment of performance standards remains an unresolved issue at this time, although guidelines are available in the literature. Subjective opinions of educators remain most popular in determining mastery levels (Popham, 1978; Meskauskas & Webster, 1975; Levine & Forman, 1973).

FINAL EVALUATION OF THE WORD IDENTIFICATION TEST BATTERY

In spring 1980, the final version of the Word Identification Test battery and the Reading Subtest of the Metropolitan Achievement Tests were administered to approximately 100 children in each grade level, one through five. The primary purpose of the study was to examine relationships between word identification skills, as measured by the various subtests in the battery, and reading comprehension, as measured by the standardized test of reading comprehension. Performance guidelines for each of the five subskills assessed in the Word Identification Test battery were determined from the results.

Method

SUBJECTS

A total of 644 pupils in grades one through five participated in the study. The pupils were from three schools in the Madison Metropolitan Public Schools, three schools in the Middleton School District, and one school in Antioch, Illinois (see Table 1). Middleton is a suburb adjacent to Madison, Wisconsin. Both Madison and Middleton are communities with a middle to upper middle class socioeconomic population. Antioch, located in the northern part of Illinois, close to the Wisconsin border, has a mixed socioeconomic population. Many residents of Antioch commute to the Chicago area or are employed in local industries.

Table 1

Subject Population by School and Grade

Grade	School districts							Totals
	Madison, Wis.			Middleton, Wis.			Antioch, Ill.	
	A	B	C	A	B	C		
1				34	43	23		100
2				30	60	17		107
3 (Structure Subtests Only)	49		44					93
3 (Phonics Subtests Only)		57		35			22	114
4	43			39			29	111
5				49	70			119
Totals	92	57	44	187	173	40	51	644

In each school, testing was conducted in intact classroom groupings. All classrooms were heterogeneous in academic ability, with pupils representing a typical range of reading skills at each grade level.

STIMULI

The Word Identification Test battery consists of five subtests within two major skill areas, phonics and structural analysis. Two of the five subtests, Consonants and Vowels, assess phonics skills; the remaining three subtests, Inflected Endings, Affixes, and Contractions & Possessives, assess structural analysis skills. All five subtests are group administered paper and pencil tests. A discussion of the criteria guiding test development, as well as a detailed description of each subtest, is presented in the Interim Report: The Refinement of the Test Battery To Assess Word Identification Skills (Johnson, Pittelman, Schwenker, & Shriberg, 1980), The Assessment of Structural Analysis Skills (Johnson, Pittelman, Schwenker, & Shriberg, 1979), and A New Approach to the Assessment of Phonics Skills (Johnson, Shriberg, Pittelman, & Schwenker, 1979).

Phonics Subtests

Consonants Subtest. The Consonants Subtest assesses 45 spelling-to-sound correspondences for single-letter consonants, consonant clusters, and consonant digraphs. Target sounds were selected according to their frequencies of occurrence in the Venezky (Note 1) tabulations of the 20,000 most common English words. All single-letter consonant and two-

letter consonant digraph correspondences with frequencies exceeding 150, and all two-letter consonant cluster correspondences with frequencies exceeding 110, were selected for inclusion in the Consonants Subtest. Table 2 presents a list of the correspondences assessed, the frequency of occurrence for each correspondence, and the position(s), initial and/or final, in which each correspondence is assessed.

The Consonants Subtest is comprised of 90 items, two items assessing each of the 45 correspondences. The subtest is designed to allow teachers to assess children's performance by category (single-letter consonants, consonant clusters, and consonant digraphs) rather than by individual correspondences. For example, a student who made errors on correspondences represented by cl-, -nt, -mp, gr-, and st-, would be viewed as having difficulty with consonant clusters in general, rather than with these five correspondences in particular.

Each item on the Consonants Subtest consists of a target synthetic word (with the target letter(s) underlined) and four response choices in picture form. The target synthetic word is a one-syllable word which conforms to regular phonological rules of the English language.

The response choices for each item are pictures whose names are well known to elementary school children. For each target item, the four response choice categories are: a Correct response choice, an Acoustically Close response choice, a Visually Close response choice, and a "Neither" (neither acoustically close nor visually close) response choice. For the six items assessing the three single-letter consonants having other

Table 2

Target Correspondences by Frequency and Position for Consonants

	Example	Total frequency	Position assessed	
			Initial	Final
Single-letter consonants				
b	bat	1,445	XX	
c (as /k/)	cup	2,433	XX	
c (as /s/)	cent	719	XX	
d	dog	2,897	X	X
f	fox	1,064	X	X
g (as /g/)	goat	722	X	X
g (as /j/ with silent e)	cage	595		XX
h	house	764	XX	
j	jug	214	XX	
k	kite	395	X	X
l	lamp	3,679	XX	
m	mice	2,711	X	X
n	sun	4,599		XX
p	pet	1,811	X	X
q (followed by u)	queen	192	XX	
r	rake	971	XX	
s (as /s/)	sink	2,171	X	X
s (as /z/)	boys	612		XX
t	tag	4,040	X	X
v (with silent e in final position)	vest	1,534	X	X
w	witch	442	XX	

Table 2 (continued)

	Example	Total frequency	Position assessed	
			Initial	Final
Consonant clusters				
stop + liquid				
br	broom	232	XX	
cl	clip	184	XX	
cr	crown	241	XX	
dr	drum	136	XX	
gr	grapes	275	XX	
pl	plug	175	XX	
pr	prince	549	XX	
tr	train	401	XX	
liquid + stop				
lt	belt	155		XX
fricative + liquid				
fl	flag	160	XX	
fr	frog	125	XX	
sl	sled	114	XX	
nasal + stop				
mp	lamp	274		XX
nd	hand	626		XX
nt	tent	1,304		XX
nasal + fricative				
nc (followed by silent e)	fence	506		XX

Table 2 (continued)

	Example	Total frequency	Position assessed	
			Initial	Final
Consonant clusters				
fricative + stop				
sc	scale	162	XX	
sp	spool	325	XX	
st	stamp	1,054	X	X
Consonant digraphs				
ch	chair	270	XX	
ng	swing	401		XX
ph	phone	188	X	X
sh	ship	427	X	X
th (voiceless)	wreath	329	X	X

common sound correspondences--c, g, and s, the Acoustically Close category was changed to be the Other Common Sound Correspondence.

When taking the test, students are directed to read the synthetic word silently to themselves and to determine the sound represented by the underlined letter(s). Next, they are told to listen as the examiner reads the names of the four picture response choices for that item. Students are then instructed to draw a circle around the picture whose name began (or ended) with the sound represented by the underlined letter(s) in the target synthetic word. Figure 1 is a copy of the directions and practice items for consonant correspondences in initial position from the Consonants Subtest.

Vowels Subtest. The Vowels Subtest assesses 29 different spelling-to-sound correspondences. Sounds selected for testing include the 5 short and 5 long vowels, 4 "other" frequently occurring single-letter vowels (i.e., single-letter vowels that corresponded to sounds that are neither short nor long), and 14 two-letter vowel clusters. Target vowels sounds were selected according to their frequencies of occurrence in the Venezky (Note 1) corpus of the 20,000 most common English words. The frequency ranges for the vowels selected for inclusion in the test, by category, are: short vowels = 1,458 to 7,554; long vowels = 508 to 1,870; "other" single-letter vowels = 117 to 243; and two-letter vowel clusters = 123 to 723.[5] A list of the vowels assessed, as well as of their frequency of

---

[5]The frequency value of 723 represents the summed frequencies of 333 for or (as in porch) and 390 for or (as in corn). Although Venezky differentiates between the two correspondences, they are treated as one correspondence in the present study.

PHONICS: Consonants
Initial Position

In each row, look at the made-up word. Notice that there is a letter or letters underlined at the beginning of that word. Read the word to yourself and decide how the underlined part sounds. Then listen carefully as I read the names of the pictures in the row and decide which picture name begins with the sound of the underlined letter or letters in the made-up word. Draw a circle around that picture.

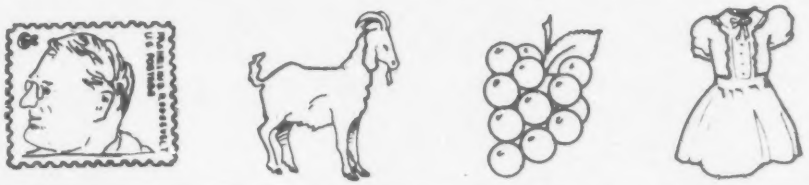
A. fape	
B. cag	
C. groon	

Figure 1. Directions and practice items from the first page of the Consonants Subtest for Initial Position.

occurrence and position(s) in words (median and/or final), is presented in Tables 3 and 4.

The Vowels Subtest is comprised of 56 items, two items assessing each of the 29 correspondences. As with the Consonants Subtest, it was intended that children's performance by analyzed by category (in terms of short vowels, long vowels, "other" single-letter vowels, and two-letter vowel clusters), rather than by individual correspondences. For example, a student who made errors on correspondences represented by ou, ai, oo, and ow, would be viewed as having difficulty with vowel clusters in general, rather than with these four correspondences in particular.

The format of the Vowels Subtest is the same as the format of the Consonants Subtest; each item on the Vowels Subtest consists of a target synthetic word (with the target letter(s) underlined) and four response choices in picture form. The target synthetic word is a one-syllable CVC or CCVC word which was constructed to conform to regular phonological rules of the Enlish language.

The response choices for each item are pictures whose names are well known to elementary school children. For each target item, the four response choice categories are: a Correct choice, an Acoustically Close response choice, a Visually Close response choice, and a "Neither" (neither acoustically close nor visually close) response choice.

As with the Consonants Subtest, students are first directed to read the synthetic word silently to themselves and to determine the sound represented by the underlined letter(s). Next, they are told to listen as the examiner reads the names of the four picture response

Table 3

Target Correspondences by Frequency and Position for

Short, Long, and Other Single-letter Vowels

	Example	Total frequency	Position assessed	
			Medial	Final
Short vowels				
a	hat	3,121	XX	
e	dress	3,241	XX	
i	fish	7,554	XX	
o	mop	1,590	XX	
u	drum	1,458	XX	
Long vowels				
a	rake	1,870	XX	
e	mete	503	XX	
i	hive	968	XX	
o	rope	1,292	XX	
u[a]	flute	967	XX	
Other single-letter vowels				
a	ball	147	XX	
o	glove	159	XX	
o	dog	117	XX	
y (sometimes followed by silent e)	fly	243		XX

[a]While the sound of u in fuse (the diphthong ju or iu) has a greater frequency than the sound of u in flute (the simple vowel u), only the latter letter-sound correspondence was used in the Vowels Subtest. This is because there are few one-syllable picturable words with u as in fuse that are well-known to elementary school pupils. Except for when the sound correspondence for the letter u occurs in initial position, the two sound correspondences above for u are considered to be very similar.

Table 4

Target Correspondences by Frequency and

Position for Vowel Clusters

	Example	Total frequency	Position assessed	
			Medial	Final
Vowel clusters				
ai	train	261	XX	
au	taut	175	XX	
ar[a]	barn	532	XX	
ay	play	142		XX
ea	seal	320	XX	
ea	bread	135	XX	
ee	feet	294	XX	
er[a]	fern	387	XX	
oo	moon	198	XX	
or[a,b]	porch	723	XX	
ou	cloud	238	XX	
ow	gown	123	XX	
ow	snow	130	X	X
ur[a]	purse	204	XX	

[a]In this subtest, all vowel + r combinations are treated as vowel clusters. The authors are aware, however, that er and ur are simple vowels, whereas ar and or are vowel + r combinations.

[b]See footnote 5 in this chapter.

choices for that item. Students are then instructed to draw a circle around the picture whose name contains the same medial (or final) vowel sound as the sound corresponding to the underlined letter(s) in the target synthetic word. Figure 2 is a copy of the directions and practice items from the Medial Vowels section of the Vowels Subtest.

Structure Subtests

Inflected Endings Subtest. The Inflected Endings Subtest assesses seven different target inflected endings. The selection of inflected endings was based on frequency information from the scope and sequence charts of selected basal reading series, and on published tests for inflected endings.

The Inflected Endings Subtest is comprised of 39 items with three to six items assessing each inflected ending. A primary factor in determining the number of test items for assessing any particular inflected ending, was the frequency of occurrence in the language for that inflected ending. Frequency information was gathered from the Ginn Lexicon Project Frequency Listing (Johnson & Baumann, 1979) (see footnote 1), and the American Heritage Word Frequency Book (Carroll et al., 1971) see footnote 2). The target inflected endings assessed in the subtest, as well as the number of items for each target inflected ending, are presented in Table 5.

The selection of the root words to be affixed was also based on frequency information from the Ginn Lexicon Project Frequency Listing

PHONICS: Vowels
Medial Position

In each row, look at the made-up word. Notice that there is a letter or letters underlined in the middle of that word. Read the word to yourself and decide how that letter or letters sound. Then listen carefully as I read the names of the pictures in the row and decide which picture name contains the sound of the underlined letter or letters in the made-up word. Draw a circle around that picture.

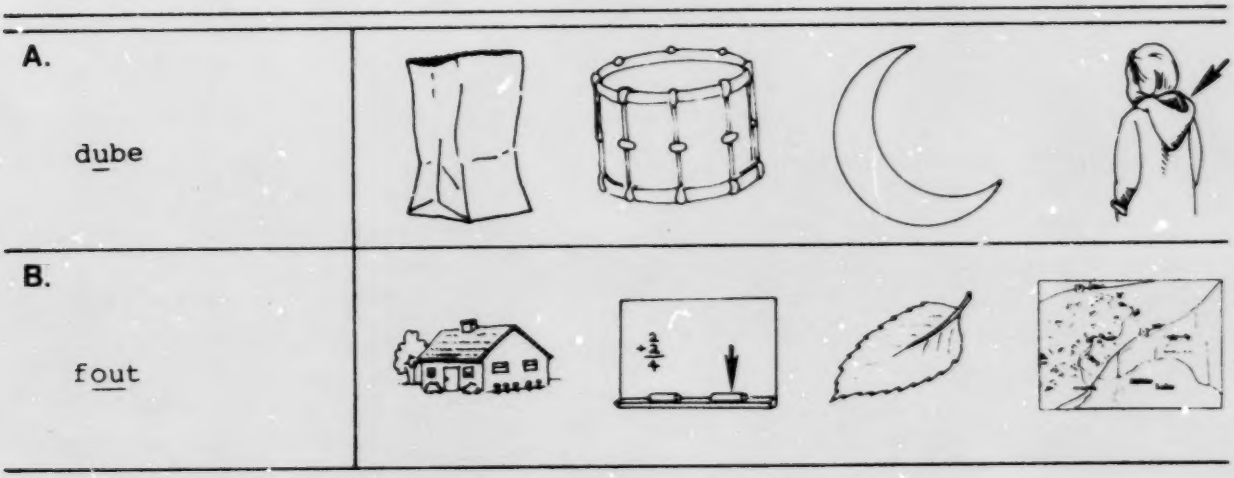


Figure 2. Directions and practice items from the first page of the Vowels Subtest for Medial Position.

Table 5

Target Inflected Endings Assessed

in the Inflected Endings Subtest

Target inflected ending	Number of items
s (plural)	5
(e)s (verb)	6
ed	6
ing	3
er	3
est	3
y	2
Other target words	
tense marker (with vowel change)	1
root (correct response)	10
Total Number of Items	39

and the American Heritage Word Frequency Book. After the root words were identified, their familiarity to young children was checked in The Living Word Vocabulary (Dale & O'Rourke, 1976) (see footnote 4).

Each item on the Inflected Endings Subtest consists of a sentence with a word missing and four response chioces beneath the sentence. Two of the four response choices are the root word with inflected endings; a third response choice is the root word alone; and the fourth response choice is the phrase "none of these."

Students are directed to read each sentence silently to themselves and to select the response choice that best completes the sentence. After circling a response choice, the children are to continue on to the next item, and so forth, until all the test items are completed. Figure 3 is a copy of the directions and practice items from the Inflected Endings Subtest.

Contractions & Possessives Subtest. The Contractions & Possessives Subtest assesses nine different forms of contracted words and the apostrophe s. The selection of contractions to be assessed was based on a review of basal materials, the Wisconsin Design for Reading Skill Development, which is a skills management system, and on frequency information from the American Heritage Word Frequency Book. Commonly taught contractions were identified and grouped into categories according to which member of the word pair was contracted. For example, contractions of will, such as I'll, we'll, he'll, and they'll formed one category. Based on frequency tabulations of contracted forms within these categories, the

Inflected Endings

Each of the sentences below has a word missing. Read each sentence to yourself. Then carefully read each of the words below the sentence. Draw a circle around the word that best completes the sentence. In some cases, "none of these" may be the best answer because the correct word is not given.

A.	Her piece of cake is _____ than mine.			
	big	biggest	bigger	none of these
B.	Doctors were _____ at the hospital.			
	need	needed	needing	none of these
C.	The puppy _____ out of the box.			
	jump	jumper	jumping	none of these

Figure 3. Directions and practice items from the first page of the Inflected Endings Subtest.

categories were then rank-ordered and a proportionate number (see Table 6) of specific contracted forms from each category were selected for assessment.

The Contractions & Possessives Subtest is comprised of 31 items, 21 items assessing contractions and 10 items assessing possessives. As in the other Structure subtests, the number of items for assessing each contracted form was based on frequency information. The target possessive and contracted forms assessed in the subtest, as well as the number of items for each form, are presented in Table 7.

Each item on the Contractions & Possessives Subtest consists of a sentence which contains an underlined contracted or possessive word and four response choices. The response choices for all the items which assess a contracted form not ending in apostrophe s consist of (1) the correct response choice; (2) and (3) two response choices consisting of two-word combinations which could make sense in the context of the sentence but which do not correspond to the contracted form; and (4) the phrase none of these. The formation of response choices for target words ending in apostrophe s was the same, regardless of whether the apostrophe s represented a possessive or a contraction. The four response choices consisted of (1) the word possessive; (2) and (3) two response choices consisting of one word of the two word phrase which comprised the target contracted form; and (4) the phrase none of these. The only exception in the format of the response choices was when none of these was the correct answer.

Table 6

Frequency of Contractions[a] and Rank Order of Contraction Categories

Contraction	Contraction frequency	Rank order by category
n't		
aren't	239	1
doesn't	590	
hasn't	98	
shouldn't	90	
weren't	176	
won't	756	
is		
it's	2,178	2
here's	118	
what's	482	
'll		
it'll	65	3
they'll	120	
you'll	524	
've		
they've	53	4
you've	317	
'd (had)		
I'd	534	5
we'd	not listed	
'd (would)		
she'd	130	6
it'd	11	
'm		
I'm	1,848	7
's (us)		
let's	892	8
're		
you're	848	9

[a]Based on information from the American Heritage Word Frequency Book.

Table 7

Target Forms in Contractions &

Possessives Subtest

Target forms		Occurrences
n't	(not)	6
'll	(will)	3
'd	(would)	2
've	(have)	2
'd	(had)	2
'm	(am)	1
's	(us)	1
're	(are)	1
's	(is)	3
's	(possessives)	10
Total		31

Students are directed to read each sentence silently and to select the response choice (provided beneath the sentence) that best completes the sentence. After circling a response choice, the children are to continue on to the next item, and so forth, until all the test items are completed. Figure 4 is a copy of the directions and practice items from the Contractions & Possessives Subtest.

Affixes Subtest. The Affixes Subtest assesses children's knowledge of 18 target affixes, 8 prefixes, and 10 suffixes. The selection of affixes for assessment on the Subtest was based on frequency information gathered from scope and sequence charts of basal reading series, published tests of affixes, and the SWRL Lexicon (see footnote 3). The Affixes Subtest has a total of 54 items with three items assessing each target affix. A list of the 18 target affixes assessed in the subtest is presented in Table 8.

The selection of the root words to be combined with the target affixes was based on two criteria: (a) a root word had to combine with at least two other affixes so that foils could be created for the test items; and (b) the familiarity of the affixed root word to second, third, and fourth grade pupils had to be as consistent as possible both within and across test items. The Living Word Vocabulary was the source used to determine the grade-level familiarity of the affixed words.

Each item of the subtest consists of a one or two line prose description of the affixed word and four response choices. In order to draw children's attention to the target root word, the root word is also

Contractions and Possessives

Read each sentence below carefully. Then decide which meaning the apostrophe mark (') has in the underlined word. Circle the choice given below the sentence that tells the meaning of the underlined word. In some cases, "none of these" may be the correct answer because the real meaning of the apostrophe in the underlined word is not given.

A.	The cuckoo clock's been broken for a long time. possessive ' clock is clock has none of these
B.	There wasn't any question that Andy was the best runner. possessive ' was none was no none of these
C.	The wind's energy is used by windmills to raise water. possessive ' wind has wind is none of these

Figure 4. Directions and practice items from the first page of the Contractions & Possessives Subtest.

Table 8

Target Affixes Assessed in the Affixes Subtest

Prefixes	Suffixes
re-	-ful
non-	-or/-er
dis-	-less
un-	-able
in-/im-	-ment
sub-	-ly
inter-	-(e)ous
pre-	-en
	-ness
	-(t)ion

printed next to the item number. The four response choices consist of the correct response and three foils of other affixed forms of the same root word which do not fit the description or definition given in the item stem.

As with the other two structure subtests, children work independently, reading the prose description for each item and then circling a response choice. As soon as a child completes an item, he or she continues on to the next item. Figure 5 is a copy of the directions and practice items from the Affixes Subtest.

Procedure

Each class participating in the study was given the designated subtests from the Word Identification Test battery and the Reading Subtest of the Metropolitan Achievement Tests. Table 9 indicates which of the five Word Identification Subtests were administered at each grade level, and the corresponding levels of the Reading Subtest of the Metropolitan Achievement Tests. As shown in the table, fifth grade students received only one subtest from the Word Identification Test battery and first grade students were given two, while second, third, and fourth grade students each received three of the give Word Identification Subtests. (Although it would have been appropriate for third grade subjects to receive all five subtests, an effort was made to limit the participation time per class.)

The subtests were administered in varying orders, depending on the sequence that best fit the time allocations specified by the schools.

Affixes

Look at each row and read the word in the small box. This word is the root or base word for the sentence. Now read the sentence. Below the sentence are four answers, each containing the root word plus another word part or parts. Draw a circle around the word that is described or defined in the sentence.

A. happy	A word that describes a person who is not happy: unhappy happiest happily unhappiness
B. sweet	A word that means the quality of being sweet: unsweetened sweetness sweeter sweetest
C. drive	A word that describes a car that can not be driven: driving driver drivable undrivable

Figure 5. Directions and practice items from the first page of the Affixes Subtest.

Table 9

Subtests by Grade Level from the Word Identification Test Battery and the Metropolitan Achievement Tests

	Phonics Subtests		Structure Subtests			Metropolitan Achievement Test (Form JI)
rade	Consonants	Vowels	Inflected Endings	Affixes	Contractions & Possessives	Reading Subtest
1	X	X				Primary I
2	X	X	X			Primary II
3						
A	X	X				Elementary
B			X	X	X	Elementary
4			X	X	X	Elementary
5				X		Intermediate

When the Phonics Subtests were administered, however, the Consonants Subtest was always given before the more difficult Vowels Subtest.

An administrator's Manual was prepared for each subtest, and the directions for administering each subtest were read from the appropriate manual. The procedure for administering each of the subtests is described below.

CONSONANTS SUBTEST

After the booklets for the Consonants Subtest were distributed and the appropriate student identification information entered on the cover, the examiner explained to the students that they would be listening for consonants sounds at the beginning of words. The examiner then worked with the students on three practice items.

For each practice item, students were directed to look at the synthetic word in the box and to pronounce the word silently to themselves. They were told to especially note the underlined letter(s) in the word, and to determine the sound of that underlined part. Next, the examiner read the names of the four pictures in the row. The students' task was to circle the picture whose name began with the same sound as the sound of the underlined part of the synthetic word.

Following discussion of the practice items (between 5 and 10 minutes), the examiner paced the children through all the test items. The examiner instructed students to, "put your finger on Row # __ and say the made-up

word to yourself." The examiner then named the four pictures in the row. This procedure was repeated for each item on the test.

When the Initial Consonants section of the Consonants Subtest was completed, the examiner explained to the students that they would next be listening for consonants sounds at the end of words. The examiner worked with students on two practice items for final consonants sounds. At the end of the practice period, the examiner led the children through the actual test items saying each picture name as before. Total test time for both the Initial and Final Consonant sections (including practice items) ranged from 30 to 55 minutes.

VOWELS SUBTEST

The Vowels Subtest was administered in the same way as the Consonants Subtest, except that students were instructed to listen to vowels sounds in the middle of words (51 items) and at the end of words (5 items). Practice items were provided for items in both positions. Total test time (including practice items) ranged from 27 to 45 minutes for the entire Vowels Subtest (both Medial and Final positions).

INFLECTED ENDINGS SUBTEST

After the booklets for the Inflected Endings Subtest were distributed and the appropriate student identification information entered on the cover, students worked with the examiner on three practice items; the

first practice item was written on the chalkboard and also appeared in the test booklet.

Students were then directed to work on the items independently until they completed all six pages of the test booklet, after which they could either draw a picture on the back cover of the booklet or do quiet seatwork. Total time for the administration of the Inflected Endings Subtest (including practice items) ranged from 17 to 36 minutes.

CONTRACTIONS & POSSESSIVES SUBTEST

The Contractions & Possessives Subtest was administered in the same way as the Inflected Endings Subtest except that pupils were told that the test booklet was about the two meanings of the apostrophe mark. As in the Inflected Endings Subtest, the children did three practice items with the examiner and then were directed to work independently on the rest of the test booklet. The only notable difference in procedure was that all three practice items from the Contractions & Possessives Subtest were written both on the chalkboard and in the test booklets. Total time for the administration of the Contractions & Possessives Subtest (including practice items) ranged from 17 to 35 minutes.

AFFIXES SUBTEST

The procedure for administering the Affixes Subtest was the same as the procedure followed for the two other structure subtests, except pupils were told that the booklet was about prefixes and suffixes--word

parts that are added to the beginning or end of a word to change its meaning or use in a sentence.

Students worked with the examiner on three practice items (none of which was written on the chalkboard) and then were directed to work indpendently on the remaining items in the booklet. Total time for the administration of the Affixes Subtest (including practice items) ranged from 23 to 35 minutes.

READING SUBTEST OF THE METROPOLITAN ACHIEVEMENT TESTS

After the booklets for the Reading Subtest of the Metropolitan Achievement Tests were distributed and the required student information entered on the cover, the examiner directed the children to open their booklets to a specific page. Following a discussion of the sample items, the examiner directed the children to work independently on the next several pages until they reached the word "STOP." When the children had completed all the items on the subtest they were permitted to do quiet seatwork of their choice. Total time for the administration of the Reading Subtest (including the sample items) took approximately 40 minutes for the first grade, and approximately 45 minutes for grades two through five.

Each of the five Word Identification Subtests, as well as the Metropolitan Reading Subtest, was administered separately, with a break between subtests. Testing was conducted in one or two sessions, depending on the number of subtests to be administered and the amount of class time

allocated. Testing sessions were usually scheduled over two separate days although, in a few cases, the sessions were broken up by the lunch period. For one school, the two sessions were one week apart. All testing took place between March and May 1980.

With the exception of the school in Antioch, Illinois, all testing was conducted by either specially trained personnel hired by the Wisconsin Research and Development Center or by Project staff. In Antioch, the tests were administered by classroom teachers under the supervision of the Reading Consultant, a former member of the Project staff.

Results

The primary purpose of the final study was to obtain word identification and reading comprehension data to establish performance guidelines for mastery on the Word Identification Test battery. In addition, summary statistics on each of the word identification subtests were calculated and are reported below.

PHONICS TEST

Summary statistics for students' performance on the two Phonics Subtests (Consonants and Vowels) are presented in Table 10. As anticipated, children performed better on the Consonants Subtest than on the Vowels Subtest.

T-tests for significant differences in performance due to sex and grade level of subjects were performed on the data. Summary information

Table 10

Summary Statistics for the Consonants and Vowels Subtests

(n = 339)

Name of subtest	Number of items	$\bar{X}$ % correct	SD	Standard error	Range % correct	Hoyt estimate of reliability
Consonants (Grades 1, 2, 3)	90	89.19	10.87	2.36	23.33 - 100.00	.94
Vowels (Grades 1, 2, 3)	56	72.35	25.40	2.81	7.14 - 98.21	.96

from these t-tests is presented in Tables 11 and 12. No difference in performance due to sex of subjects was observed on either of the two subtests. Grade level differences, on the other hand, were significant for both the Consonants and Vowels Subtests with one exception; the performance of second and third grade subjects was not significantly different on the Consonants Subtest. This may be because by the end of second grade most students have mastered the major consonants correspondences. This suggestion is supported by the high mean score (92.11%) obtained by second graders. At higher grade levels, therefore, relatively small changes in consonant scores are observed.

In contrast, significant grade level increases in scores assessing the vowel correspondences are apparent on the Vowels Subtest. Besides the expected difference in performance of second grade over first grade, third grade performance shows a significant increase over second grade performance. Evidence from previous studies suggests that fourth grade students would achieve still higher scores on vowels correspondences, although the differences might not be statistically significant.

Subjects' performance on the Phonics Subtests was also examined by item categories within each subtest. This information, as well as reliability estimates for the subtests, is presented in Tables 13 and 14 for Consonants and Vowels.

The rank-ordered listing of mean percent correct scores by consonant category differs somewhat from results of the previous study: Children

Table 11

Summary of t-tests for Differences

Due to Sex

Name of subtest	Sex	N	$\bar{X}$ % correct	DF	t-value	Probability
Consonants	Boys	171	69.73	337	.15	.878
	Girls	168	69.85			
Vowels	Boys	160	73.02	314	.07	.945
	Girls	156	72.87			

Table 12

Summary of t-tests for Differences

Due to Grade Level

Name of subtest	Grade Level	N	$\bar{X}$ % correct	DF	t-value	Probability
	1	119	81.16			
				223	6.58	.000
Consonants	2	106	92.11			
				218	.69	.494
	3	114	94.31			
	1	99	48.85			
				203	11.51	.000
Vowels	2	106	73.47			
				215	3.06	.003
	3	111	79.25			

Table 13

Summary Statistics for Consonants Subtest

by Item Category and Grade Level

(n = 339)

Item category	Number of items	Mean percent correct	
Single-letter consonants	36	Grade 1	70.10
		Grade 2	71.15
		Grade 3	71.22
Consonant clusters	38	Grade 1	66.39
		Grade 2	75.15
		Grade 3	74.56
Consonant digraphs	10	Grade 1	70.67
		Grade 2	81.32
		Grade 3	83.33
Variant single-letter consonants (c, g, and s)	6	Grade 1	21.43
		Grade 2	41.98
		Grade 3	46.93
Total test	90		

Table 14

Summary Statistics for Vowels Subtest

by Item Category and Grade Level

Item category	Number of items	Mean percent correct	
Long vowels	10	Grade 1	44.34
		Grade 2	71.98
		Grade 3	73.87
Short vowels	10	Grade 1	49.49
		Grade 2	73.96
		Grade 3	77.48
Vowel clusters	28	Grade 1	64.47
		Grade 2	89.86
		Grade 3	98.13
Other single-letter vowels	8	Grade 1	33.71
		Grade 2	53.77
		Grade 3	59.23
Total test	56		

in both studies performed poorest on variant single-letter consonants (i.e., those consonants letters--c, g, and s--having more than one common sound correspondence). In the present study, performance was highest on the consonant digraph category rather than on the single-letter consonants category which had shown highest performance in the previous study. The second and third grade subjects in the present study performed considerably better on the items assessing digraphs than they did on items assessing single-letter consonants. Whether this was due to more recent instruction on digraphs or to chance was not determined. Overall, for the present subtest, mean percent correct on consonant categories ranged from 78.44 for digraphs to 36.78 for variant single-letter consonants.

On the Vowels Subtest, performance on categories followed an identical pattern to that observed in the previous study: Children did best on vowel clusters and least well on variant vowels correspondences (vowel letters which correspond to a number of frequently occurring sounds). Mean percent correct scores ranged from 84.15 for items assessing vowel clusters, to 48.90 for the variant other single-letter vowel category. Although performance increases with grade level across all categories, the most notable feature of the data is the very large jump in mean performance on all item categories between first and second grades (see Table 14). Although performance for all categories is yet higher for third grade, the mean differences are not as great. Mastery of vowels correspondences appears to occur somewhat later for most students than does the mastery of consonants correspondences.

As discussed earlier, the Phonics Subtests were designed to facilitate the analysis of error patterns--response foils were carefully created using speech production and confusion matrix information for each subtest item. The item analysis data were tallied item-by-item in order to obtain selection rates for subjects' errors. These error selection rates are presented by grade in Table 15 for the Consonants Subtest and in Table 16 for the Vowels Subtest.

For the Consonants Subtest, the foil categories were Acoustically Close (Acoustic), Visually Close (Visual), the other frequently occurring sound of a variant consonant (Other), or Neither Acoustically Close nor Visually Close to the target letter(s) or sound (Neither). Rates of selection for these foil categories were consistent across all three grade levels: "Other" foils had the highest rate of selection, and "Neither" foils had the lowest rate. The very low rate of selection for "Neither" foils across all three grade levels is interesting. A selection rate of approximately 25% for this foil type would indicate a pattern of random guessing. As in the previous study, subjects at all grade levels tended to be strategic, rather than random, in their selection of responses.

For the Vowels Subtest, the foil categories were Acoustically Close (Acoustic), Visually Close--either one of the two letters for a target cluster or digraph (1 of 2), another common sound correspondence for single-letter vowels (Other), and Neither Acoustically Close nor Visually Close to the target letter(s) or sound (Neither). As with the Consonants

Table 15

Percent of Times Each Consonant Foil Was Selected

Grade	Visual (present in 90 items)	Acoustic (present in 79 items)	OR	Other (present in 11 items)	Neither (present in 90 items)
1	7.26%	4.26%		46.9%	1.3%
2	2.17%	1.28%		36.9%	0.33%
3	1.54%	0.91%		31.2%	0.22%

Note. The percentage figure indicates the number of times the foil was incorrectly selected.

Table 16

Percent of Times Each Vowel Foil Was Selected

Grade	Acoustic (present in 56 items)	Visual: Other (present in 25 items)	Visual: OR	Visual: 1 of 2 (present in 31 items)	Neither (present in 56 items)
1	11.23%	31.84%		17.35%	4.27%
2	8.20%	24.2%		12.3%	1.84%
3	6.69%	22.16%		8.61%	1.25%

Note. The percentage figure indicates the number of times the foil was incorrectly selected.

Subtest, the rates of selection for the various error categories in the Vowels Subtest were completely consistent across all three grade levels tested. The highest rate of selection was observed for the "Other" category, and the lowest rate was for the "Neither" category.

Because error analyses are generally done on the responses of individual students, rather than tallied for groups, five second grade subjects were randomly selected from each of three score ranges for the Vowels and Consonants Subtests and their errors were tallied. Students scoring one standard deviation above and below the mean comprised the High and the Low groups, respectively. Students scoring at or very near the mean comprised the average group. The students' errors were tallied individually and then summed by error type.

A breakdown of children's error patterns for the Consonants Subtest are presented in Table 17. The patterns do not differ markedly from those observed from the entire sample: the "Other" response foil had the highest percentage of selection, whereas the "Neither" response foil had the lowest. A notable observation is that high scoring subjects making errors did not select the visual response foils. The fact that average and low scoring subjects frequently selected the visual foil may indicate a tendency for visual matching of letters--that is, average and low scoring subjects may have mentally spelled out the picture names and then proceded to "match" the letter(s) of the picture name to the letter(s) of the target synthetic word.

The findings were similar for the Vowels Subtest: the "Other" response foil was most frequently chosen by subjects in all three score

Table 17

Number of Times Consonant Foils Were Selected

by Each Score Group

(n = 5 per group)

Group	Visual present in 90 items (number possible responses = 450)	Acoustic present in 79 items (number possible responses = 395)	OR	Other present in 11 items (number possible responses = 55)	Neither present in 90 items (number possible responses = 450)	Total number of errors
High	0 (0%)	1 (0.2%)		16 (29.1%)	0 (0%)	17
Average	33 (7.3%)	18 (4.5%)		24 (43.6%)	4 (0.9%)	79
Low	71 (15.8%)	37 (9.4%)		18 (32.7%)	26 (5.8%)	152

Note. The percentage figure in parentheses indicates the percentage of times the foil was incorrectly selected.

ranges, whereas the "Neither" response foil was selected the least (see Table 18). Most of the errors made were on items assessing the long or short sound of a single-letter vowel. In such items, the "Other" response foil represented the long or short pronunciation counterpart of the target vowel sound (e.g., long o, if short o were the correct answer). One interpretation as to why children made so many errors on this type of item may be that many of these subjects had not yet learned that the silent e is a marker for long vowels. As expected few subjects made random guesses (as indicated by the low selection of the "Neither" response foil); the percentage of selections of the "Neither" response foil was greatest for children who performed the poorest on the Vowels Subtest.

Structure Subtests

Summary statistics for the three Structure Subtests (Inflected Endings, Contractions & Possessives, and Affixes) are presented in Table 19. As indicated in the table, children performed best on the Inflected Endings Subtest. The high performance on this subtest was anticipated, because Inflected Endings is one of the earliest components of structural analysis to be taught in most reading programs.

T-tests for significant differences in performance due to sex and grade level of subjects were also performed for each subtest. Summary information from these t-tests is presented in Tables 20 and 21, respectively. The only case in which a significant difference due to sex was observed was in the Inflected Endings Subtest (which was administered to second, third, and fourth grade subjects): Girls outperformed boys

Table 18

Number of Times Vowel Foils Were Selected

by Each Score Group

(n = 5 per group)

Group	Acoustic	Visual			Neither	Total number of errors
		Other	OR	1 of 2		
	present in 56 items (number possible responses = 280)	present in 25 items (number possible responses = 125)		present in 31 items (number possible responses = 155)	present in 56 items (number possible responses = 280)	
High	5 (1.8%)	7 (5.6%)		3 (1.9%)	0 (0%)	15
Average	21 (7.5%)	48 (38.4%)		25 (16.1%)	7 (2.5%)	101
Low	41 (14.6%)	47 (37.6%)		39 (25.2%)	22 (7.8%)	149

Note. The percentage figure in parentheses indicates the percentage of times the foil was incorrectly selected.

Table 19

Summary Statistics for Structure Subtests

Name of subtest	N	Number of items	$\bar{X}$ % correct	SD	Standard error	Range % correct	Hoyt estimate of reliability
Inflected Endings (Grades 2, 3, 4)	306	39	88.90	14.34	1.65	20.51 - 100	.91
Contractions & Possessives (Grades 3, 4)	201	31	83.89	15.55	1.77	35.48 - 100	.86
Affixes (Grades 2, 3, 4)	319	54	83.40	12.88	2.24	22.22 - 100	.84

Table 20

Summary of t-tests for Differences Due to Sex on Structure Subtests

Name of subtest	Sex	N	$\bar{X}$ % correct	df	t-value	Probability
Inflected Endings	Boys	137	86.36	304	2.82	.005
	Girls	169	90.96			
Contractions & Possessives	Boys	87	81.27	199	.74	.500
	Girls	114	82.92			
Affixes	Boys	142	83.09	317	.39	.694
	Girls	177	83.66			

Table 21

Summary of t-tests for Differences Due to Grade Level on Structure Subtests

Name of subtest	Grade level	N	$\bar{X}$ % correct	df	t-value	Probability
Inflected Endings	2	106	79.41	194	5.86	.000
	3	90	92.28	198	2.74	.007
	4	110	95.27			
Contractions & Possessives	3	92	80.79	199	2.64	.009
	4	109	86.50			
Affixes	3	93	77.26	198	3.38	.001
	4	107	83.99	224	2.82	.005
	5	119	87.68			

with a resultant t-value of 2.82 with probability = .005. This finding was not surprising since girls generally outperform boys in reading tasks in the primary grades. All grade level differences were significant. It is interesting to note the large jump in performance between second and third grade subjects on the Inflected Endings Subtest (Table 21) in contrast to the minimal rise in mean scores between third and fourth grade subjects on this subtest. This may be attributed to the fact that many inflected endings are taught in third grade.

Another purpose of the study was to confirm the reliability of test items. An item analysis indicated that, for all three Structure Subtests, each individual item showed reliability with the total subtest; hence, revisions on the subtests are not needed. (Reliability estimates for each subtest are presented in Table 19.) Students' performance on each item category within each of the three Structure Subtests was also examined. This performance information is presented in Tables 22, 23, and 24 for Inflected Endings, Contractions & Possessives, and Affixes, respectively.

In general, the rank-ordered listing of mean percent correct by category within the Inflected Endings Subtest confirms results of the previous study: Children performed best on items assessing the ing endings (mean percent correct = 97.9) and, excluding items in which the correct response choice was "none of these," least well on items assessing the tense marker (mean percent correct = 80.4). In the previous administration of the Inflected Endings Subtest, performance was generally poorest on items with a correct response of "none of these." As indicated in Table 22, this was also true in the current study.

Table 22

Rank-ordered Listing of Mean Percent Correct

on Item Categories in Inflected Endings Subtest

(n = 306)

Item category	Number of items	Mean percent correct
ing	3	97.9
root (correct response)	8	92.8
(e)s (plural)	4	92.5
y	2	92.3
(i)ed	5	90.8
er	3	90.8
est	2	88.7
(e)s (verb)	5	85.5
tense marker (with vowel change)	1	80.4
"none of these" (correct response)	6	77.4
Total test	39	77.4

Table 23

Rank-ordered Listing of Mean Percent Correct

on Item Categories in Contractions & Possessives Subtest

(n = 201)

Item categories	Number of items	Mean percent correct
am	1	98.5
is	3	96.3
will	2	95.5
are	1	95.0
have	1	95.0
would	2	83.1
not	3	82.6
"none of these" (correct response)	7	81.2
possessives	8	79.7
us	1	75.6
had	2	68.2
Total test	31	83.9

Table 24

Rank-ordered Listing of Mean Percent Correct

on Item Categories in Affixes Subtest

(n = 319)

Item categories	Number of items	Mean percent correct
Prefixes		
non-	3	95.7
re-	3	93.7
un-	3	93.7
dis-	3	87.6
in-/im-	3	87.0
sub-	3	78.2
pre-	3	70.9
inter-	3	52.2
Suffixes		
-ful	3	93.7
-able/-ible	3	93.5
-less	3	92.5
-or/-er	3	91.8
-ment	3	90.2
-en	3	88.0
-ly	3	83.2
-ness	3	79.2
-(i)(e)ous	3	70.0
-(t)ion	3	59.9
Total test	54	83.4

A similar pattern of poor performance on items with the correct response "none of these" also had been noted in the previous administration of the Contractions & Possessives Subtest. To evaluate the impact of the "none of these" foil on the Contractions & Possessives Subtest, revisions were made in the response choices of four items prior to the final administration of the test in the current study: A response choice in one item having the correct answer as "none of these" was changed so that the correct contracted form was present; the response choices for two items in which the correct contracted form was present were changed so that the foils "none of these" became the correct answers; and the correct response choice for one possessive item was changed to "none of these" (formerly, there were no possessive items with a correct response of "none of these").

A comparison of performance on these items on the two test versions indicates that children do perform differently when the correct answer is not present (i.e., when "none of these" is the correct response choice), as opposed to when the correct answer is among the response choices. Because of this effect, items with "none of these" as the correct response are grouped as a separate category in Tables 22 and 23 rather than being incorporated into the appropriate item category. (The response choice "none of these" was not present in the Affixes Subtest.)

As in the previous study on the Contractions & Possessives Subtest, children performed best on the contracted forms of _am_ and _is_. Mean percent correct scores on the contraction categories ranged from 98.5 (for _am_) and 96.3 (for _is_) to 68.2 (for _had_).

Overall performance on the prefixes and suffixes categories in the Affixes Subtest varied according to grade level. As indicated in Table 25, third grade students performed slightly better on suffices than on prefixes. This finding is in line with the result of the previous study in which second and third grade students performed better on items assessing suffixes than on items assessing prefixes. Notably, however, fourth and fifth grade subjects performed slightly better on prefixes than on suffixes.

In the rank-ordered breakdown of items within the prefix category, students across all three grade levels performed best on the prefix non- (mean percent score = 95.76), and poorest on the prefix inter- (mean percent score = 52.2). Relative performance on items within the prefix category differed from the previous study as a result of extensive revisions made for the present version of the subtest. Within the suffix category, performance in the current study was best on the suffix -ful (mean percent score = 93.7), and poorest on -(t)ion (mean percent score = 59.9). Although relative performance on items for these two suffixes was the same as in the previous study, the rank orderings of the other suffix items varied considerably because of the revisions made for the present test.

Overall performance on the Affixes Subtest was considerably higher for all categories than was observed in the previous study. This was attributed to the fact that in the current study, the test was administered to older students.

Table 25

Performance on Prefixes and Suffixes by Grade Level

Grade	Number of items	Mean percent correct	
		Prefixes	Suffixes
3	93	75.81	76.06
4	107	82.90	82.34
5	119	87.08	85.77

In summary, total group performance on all three Structure Subtests was generally similar to total group performance on the previous version. Exact rankings of specific categories varied, however, due to differences in many of the test items and the ages of the subjects in the two studies. In the previous study, all subtests were given to second and third grade students. In the current study, it appeared most appropriate to give the Inflected Endings Subtest to second, third, and fourth grade subjects; the Contractions & Possessives Subtest to third and fourth grade subjects; and the Affixes Subtest to third, fourth, and fifth grade subjects.

THE ESTABLISHMENT OF PERFORMANCE GUIDELINES FOR THE WORD IDENTIFICATION TEST BATTERY

As the review of mastery learning theory indicated, most criterion-referenced assessment devices discriminate between children who have mastered specific subskills and those who require additional instruction and practice. Typically, cutoff scores or mastery levels have been set by publishers of tests and tend to be absolute. For example, in many skills management programs, a score of 80% or better indicates mastery of a skill. Recently, however, Terwilliger (1979) pointed to a need for flexibility in setting performance standards, because the arbitrary designation of one uniform standard for all tests has not been empirically justified. This section discusses the development of performance guidelines for the Word Identification Test battery in relation to issues of skills mastery. These guidelines reflect the concern expressed by Terwilliger for flexibility in setting mastery standards. In addition to the data obtained on specific subtests, the criteria for mastery take into account childrens' reading comprehension performance and grade level.

The establishment of the performance guidelines for the Word Identification Test battery was based on an innovative, two-stage process. First, students participating in the final administration of the Word Identification Test were stratified into three comprehension ability groups: low, average, and high. These ability groups were formed on the basis of childrens' scores on the Reading Subtest of the Metropolitan

Achievement Tests, a standardized measure of reading comprehension. Children in the low comprehension group scored below their respective grade level, children in the average group scored at grade level, and children in the high group scored above grade level. The concept underlying this technique is Glaser's (1963) assertion that a criterion-referenced test is based on the notion of a continuum of knowledge acquisition ranging from no skill to perfect performance.

Second, mean scores were calculated for the three comprehension groups at each grade level for every subtest in the Word Identification Test battery. These subtest standards are presented in Table 26. Because the individual subtest scores are composites of the categories of items comprising them, additional performance standards were calculated separately for all categories within each subtest. These standards for subskill categories are listed in Tables 27 and 28 for the Phonics and Structure Components, respectively.

The rationale for including performance guidelines for subskill categories on the subtests is to provide teachers with information about areas of both strength and weakness within each subtest. Typically, a student receives one score for an entire subtest. But, of what use to a teacher is a score of 50%? While such a score may indicate the need for more practice, it does not provide the teacher with information as to the particular area of weakness. Standards for subskill categories, on the other hand, enable a teacher to use test information diagnostically. For example, with specific scores on subskill categories within the

Table 26

Performance Standards for Low, Average, and High Comprehenders on Subtests in the Word Identification Test Battery

Subtest	Grade	Comprehension Group		
		Low	Average	High
Phonics				
Consonants	1	51.2	60.7	62.9
	2	62.7	65.6	69.1
	3	63.5	68.6	69.8
Vowels	1	40.1	47.4	58.9
	2	61.2	69.3	75.9
	3	54.2	71.2	80.8
Structure				
Inflected Endings	2	49.7	70.5	89.0
	3	60.0	87.2	95.1
	4	89.4	93.2	96.4
Contractions & Possessives	3	49.1	61.8	84.1
	4	58.9	77.7	90.7
Affixes	3	46.9	54.1	81.1
	4	61.2	73.6	87.1
	5	71.5	81.6	88.3

Table 27

Performance Standards for Low, Average, and High Comprehenders on Phonics Subskills

Phonics subskills	Grade	Comprehension group		
		Low	Average	High
Consonants				
Single-letter	1	67.8	71.6	72.3
	2	72.4	70.4	71.1
	3	68.1	71.0	71.7
Clusters	1	59.4	70.5	73.3
	2	73.7	74.6	75.7
	3	70.0	75.1	75.0
Digraphs	1	62.6	78.0	76.4
	2	75.0	80.0	83.1
	3	76.0	82.4	84.5
Other (variants)	1	15.1	22.8	29.6
	2	29.8	37.3	46.3
	3	40.0	46.0	48.0
Vowels				
Long	1	35.1	44.4	56.4
	2	60.7	70.8	74.8
	3	53.3	65.7	78.3
Short	1	41.7	44.4	64.2
	2	66.4	71.2	76.6
	3	55.6	73.8	80.9
Clusters	1	54.4	65.7	76.6
	2	71.2	84.6	95.7
	3	71.8	95.4	99.9
Other single-letter	1	29.3	35.2	38.3
	2	46.4	50.5	56.5
	3	36.1	50.0	64.2

Table 28

Performance Standards for Low, Average, and

High Comprehenders on Structure Subskills

Structure subskills	Grade	Comprehension group		
		Low	Average	High
Inflected Endings				
	2	49.3	70.5	89.0
	3	60.0	87.2	95.1
	4	89.4	93.2	96.4
Contractions & Possessives				
Contractions	3	63.9	71.4	86.9
	4	70.2	81.6	91.0
Possessives	3	34.3	52.2	81.2
	4	47.5	73.8	90.4
Affixes				
Prefixes	3	47.6	53.7	80.9
	4	63.7	72.7	87.5
	5	72.9	79.9	89.4
Suffixes	3	46.2	54.4	81.3
	4	58.6	74.4	86.7
	5	70.0	83.2	87.1

Consonants Subtest, a teacher would know whether a low score could be attributed to a students' weakness in single-letter consonants, consonant clusters, consonant digraphs, or in the variant single-letter consonants and could plan instruction accordingly.

Because the data reported here is based on a single study group, there are several instances where performance between comprehension groups does not change in the expected direction. For example, for the single-letters category of the Consonants Subtest, mean scores for the three ability groups all cluster at about 70%, with the low group performing slightly better than the average and high comprehension groups. Similarly, for Consonant Digraphs, the average group obtained a slightly higher score than the high comprehension group. While a larger or different population of subjects might have produced scores in the expected direction, these discrepancies point out an important facet of skills acquisition--that is, mastery of specific subskills is not always consistently correlated with comprehension ability. This interpretation is supported by the relatively low correlations between subskill performance and comprehension ability on the two consonants categories reported above (see Table 29 for correlations listings). These data suggest the need for teachers to be flexible in judging score profiles of individual readers.

The performance guidelines for the Structure Subtests presented in Table 28 show relatively large variations between grade levels and between comprehension groups. This variation is not surprising, because structural analysis skills are taught throughout the middle elementary

Table 29

Pearson Correlations of Subtest Category Score with Metropolitan Comprehension Scores

Subtest categories	r	p
Consonants (N = 339)		
Single-letter	.2765	.000
Clusters	.4248	.000
Digraphs	.3674	.000
Other (variants)	.4656	.000
Vowels (N = 316)		
Long	.4868	.000
Short	.3961	.000
Clusters	.5715	.000
Other single-letter	.3630	.000
Inflected Endings (N = 306)	.6850	.000
Contractions & Possessives (N = 201)		
Contractions	.5794	.000
Possessives	.5381	.000
Affixes (N = 319)		
Prefixes	.5502	.000
Suffixes	.5997	.000

school years and, therefore, students would acquire mastery at various stages during this period.

Most criterion-referenced tests and skills management systems argue for across-the-board mastery cutoff scores of 80 or even 90%. Of the 42 performance standards reported in Table 26, 29 were below 80% and 38 were below 90%. The range of performance was 40.1% correct to 96.4% correct. Clearly, at times a score of 45% may be satisfactory for some students. At other times, nothing less than a score of 95% would be satisfactory.

From the inspection of performance guidelines for mastery between the various structure subskills, it is apparent that applying a uniform standard of 80, 70, or even 60% would not be appropriate. In the Contractions & Possessives Subtest, for example, third graders in the average comprehension group obtained a mean score of 71.4% on Contractions, but only 52.2% on Possessives. Hence, despite the fact that these children comprehend at grade level, they would not be regarded as "masters" of possessive forms according to traditional skills management standards. The issue to address is not whether students who score 52% on Possessives are masters of that subskill, but whether instruction aimed at raising these scores to some absolute standard will enable students to comprehend better. For this reason, the standards presented in this report are labeled "performance guidelines," rather than "cutoff scores."

According to Popham (1978), there is no true and definitive cutoff score and, therefore, the need for further practice must be based on

subjective decisions. Kriewall (1972) argues for the use of proficiency distributions. This idea is taken further by Block (1972, 1973), Millman (1973), and Terwilliger (1979), who advocate flexible mastery levels. Millman suggests that data from "masters" be used to establish standards and proposes that the cutoff scores vary so fundamental skills have higher cutoff points than nonessential skills.

The suggestions of these researchers influenced the development of the performance guidelines presented in Tables 26, 27, and 28. Because the ultimate goal of reading instruction is reading comprehension, three groups representative of the continuum of proficiency in global comprehension were used to establish flexible cutoff scores. By including the correlation listings (Table 29) in the mastery decision process, the teacher can be reasonably sure that when a low-scoring student is assigned to further practice on a skill, the skill is considered essential to comprehension.

Performance guidelines for the subskills assessed in the Word Identification Test battery are intended for use with Table 29. These correlations were computed for each subtest category with scores from the Reading Subtest of the Metropolitan Achievement Tests. Correlation information provides an additional source to help teachers make sound judgments about particular subskill scores. For example, if a second-grade student of high comprehension ability scores 48% (which is below the performance standard) on vowel items assessing "other single letter" vowel correspondences, the teacher can consult the list of correlations

(Table 29). The Pearson correlation of "other single letter" vowels with comprehension is only .3630. Therefore, rather than automatically assigning the child to further practice on these correspondences (i.e., the o in love), the teacher might decide that other uses of the student's time at this stage are more beneficial to the child's overall growth in reading skill.

CONCLUSIONS

The Word Identification Test battery was designed with attention to the major issues pertaining to skills mastery and assessment that are raised in the review of mastery learning. There were five important areas of concern in the development of the battery: (a) the basis on which target skills would be selected for inclusion; (b) the facilitation of error analysis by creating categorical distractors; (c) the ease and efficiency of test administration; (d) the independence of the test battery from any published set of materials to lessen the likelihood of teachers teaching to the tests; and (e) the establishment of flexible standards for skills mastery (performance guidelines) based on a global measure of comprehension, rather than on arbitrary cutoff scores. The manner in which each of these issues was resolved in developing the Word Identification Test battery is summarized below.

First, all five subtests comprising the battery were developed in accordance with the particular subskills that the widely used reading programs teach. Only the most frequently occurring elements of language (based on frequency data) were selected as targets for assessment. Details of the subtest specifications are documented in the Interim Report: The Refinement of the Test Battery to Assess Word Identification Skills, (Johnson, Pittelman, Schwenker, & Shriberg, 1980). Hence, the target items selected for assessment were ecologically based.

Second, the battery facilitates teachers' use of error analysis information by its use of distractor categories. Using the key to distractor

categories on the Phonics Test, for example, a teacher can analyze the error patterns of individual students with regard to visual confusion, auditory confusion, and random guessing.

Third, the battery was developed to ease the burden of classroom evaluation. Because more testing time generally means less teaching time, the subtests were designed for efficiency and for group administration. The tests are accompanied by administration manuals which include clear directions, illustrative examples, and practice items to prepare students for each test. Each of the subtests is administered separately.

Fourth, the subtests were developed independent of any particular set of materials used for teaching reading. Therefore, the likelihood of teachers teaching to the test was minimized. This is a common problem in schools where a single skills management series is used for both assessment and instruction.

Finally, the battery provides flexible standards for skill mastery. Based on a student's level of comprehension, a teacher can select the appropriate criterion score at the student's grade level for each subtest and for each skills category in the battery.

The Word Identification Test battery is a valid and reliable instrument. Although it is easy to administer, all subtests in the battery have considerable scope. The subtests enable educators to make accurate diagnostic decisions about the apportionment of instructional time on the most frequently occurring phonics and structural elements. Performance standards are provided for each subtest in the battery. The performance guidelines range from 34.3 to 96.4%, depending on the sub-

skill being measured, and the grade level and comprehension ability of the student. More eloquently than any argument appearing in the literature, this range of expected performance demonstrates the inappropriateness of arbitrarily established, rigid mastery scores. The establishment of empirically based performance guidelines for the Word Identification Test battery, on the other hand, represents a flexible and innovative solution to the issue of skills mastery.

Reference Notes

1. Venezky, R. L. A computer program for deriving spelling to sound correlations. Master's Thesis, Cornell University, 1962. Published in part in H. Levin et al., A basic research program on reading. Ithaca, NY: Cornell University, 1963.

2. Massaro, D. W. Personal communication, 1979.

References

Aaron, I. E., et al. Reading unlimited. Glenview, Ill.: Scott, Foresman, & Co., 1976.

Aaronson, S. Vocabulary instruction: Challenge of the seventies. Paper presented at the National Reading Conference, Tampa, December 1971.

Agnew, D. C. The effect of varied amounts of phonetic training on primary reading. Duke University Research Studies in Education, No. 5. Durham, N. C.: The Duke University Press, 1939.

Anderson, L. W., Scott, C. C., & Hutlock, J., Jr. The effects of a mastery learning program on selected cognitive, affective and ecological variables in grades 1 through 6. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, 1976.

Bagford, J. Management systems and comprehension do not mix! Language Arts, 1977, 54(5), 517-520.

Barber, C. Training principals and teachers for mastery learning. Educational Leadership, 1979, 37(2), 126-127.

Berk, R. A. Domain-referenced versus mastery conceptualization of criterion-referenced measurement: A clarification. Paper presented at the annual meeting of the American Educational Research Association, Boston, April 1980.

Block, J. H. Mastery learning: Theory and practice. New York: Holt, Rinehart, & Winston, Inc., 1971.

Block, J. H. Student learning and the setting of mastery performance standards. Educational Horizons, 1972, 50, 183-191.

Block, J. H. Mastery learning in the classroom: An overview of recent research. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, February 28, 1973.

Block, J. H. Schools, society and mastery learning. New York: Holt, Rinehart, & Winston, 1974.

Block, J. H. Mastery learning: The current state of the craft. Journal of Educational Leadership, 1979, 37(2), 114-117.

Block, J. H., & Burns, R. Mastery learning. In L. Shiman (Ed.), Review of Research in Education, 1977, 4, 3-49.

Blohm, P. J. Effects of skill-based instruction and skill mastery testing on general reading comprehension of tenth grade poor comprehenders; parts 1 and 2 (Technical Report No. 449). Madison: Wisconsin Research and Development Center for Individualized Schooling, University of Wisconsin, 1978. [ERIC Document Reproduction Service No. ED 153 187]

Bloom, B. S. Learning for mastery. Evaluation Comment, 1968, 1(2).

Bloom, B. S. Mastery learning. In J. R. Block (Ed.), Mastery learning: Theory and practice. New York: Holt, Rinehart, & Winston, Inc., 1971.

Bloom, B. S. Human characteristics and school learning. New York: McGraw-Hill, 1976.

Bloom, B. S. New views of the learner: Implications for instruction and curriculum. Educational Leadership, 1978, 35, 563-576.

Bloom, B. S. The new direction in educational research, alterable variables. Phi Delta Kappan, 1980, 61(6), 382-385.

Bond, G. L., Balow, B., & Hoyt, C. J. The silent reading diagnostic tests. Chicago: Lyons & Carnahan, Division of the Rand McNally College Publishing Co., 1970.

Bond, G. Y., & Wagner, E. B. Teaching the child to read. London: The MacMillan Company, Collier-MacMillan Limited, 1969.

Botel, M. Botel reading inventory. Chicago: Follett Publishing Co., 1961.

Bouma, H. Visual recognition of isolated lower-case letters. Vision Research, 1971, 11, 459-474.

Breen, L. C. Vocabulary development by teaching prefixes, suffixes, and root derivatives. The Reading Teacher, 1960, 14(2), 93-97.

Broska, K. F., Hodges, K. T., Patrick, F. J., Williams, G. M., & Oseroff, A. Clinical teacher desired pupil behaviors: An individualized reading curriculum. A publication of the Clinical Teacher Model Special Project. Tallahassee: Florida State University, 1973. [ERIC Document Reproduction Service No. ED 082 439]

Brown, G. M., & Cottrell, A. B. California phonics survey. Monterey, Calif.: McGraw-Hill, 1963.

Bruner, J. The process of education. Cambridge: Harvard University Press, 1960.

Burns, R. B. Mastery learning: Does it work? Journal of Educational Leadership, 1979, 37(2), 110-114.

California achievement tests. Monterey, Calif.: McGraw-Hill, 1977.

Carroll, J. B. A model of school learning. Teachers College Record, 1963, 64, 723-733.

Carroll, J. B. Problems of measurement related to the concept of learning for mastery. Educational Horizons, 1970, 48(3), 71-80.

Carroll, J. B. Problems of measurement related to the concept of learning for mastery. In J. H. Block (Ed.), Mastery learning theory and practice. New York: Holt, Rinehart, & Winston, 1971.

Carroll, J. B., Davies, P., & Richman, B. The American heritage word frequency book. New York: Houghton-Mifflin, 1971.

Cassidy, J. High school graduation requirements. Journal of Reading, 1978, 21(5), 398-401.

Chomsky, N., & Hall, M. The sound pattern of English. New York: Harper & Row, 1968.

Clymer, T. The utility of phonic generalizations in the primary grades. The Reading Teacher, 1963, 16, 252-258.

Clymer, T., et al. Reading 720. Lexington, Mass.: Ginn & Company, 1976.

Clymer, T., et al. Reading 720, Rainbow Edition. Lexington, Mass.: Ginn & Company, 1979.

Currier, L. B. Phonics and no phonics. Elementary School Journal, 1923, 23, 448-452.

Currier, L. B., & Duguid, O. C. Phonics or no phonics? Elementary School Journal, 1916, 17, 286-287.

Dale, E., & O'Rourke, J. The living word vocabulary. Elgin, Ill.: Dome Inc., 1976.

Deighton, L. C. Vocabulary development in the classroom. New York: Columbia University Press, 1959.

Doren, M. Doren diagnostic reading test of word recognition skills. Circle Pines, Minn.: American Guidance Service, 1973.

Duffy, G. G. Maintaining a balance in objective-based reading instruction. The Reading Teacher, 1978, 31(5), 519-523.

Duffy, G. G., & Sherman, G. B. Systematic reading instruction (2nd. ed.). New York: Harper & Row, 1977.

Durrell, D. Improving reading instruction. Chicago: Harcourt, Brace and World, 1956.

Durkin, D., & Meshover, L. Phonics knowledge survey. New York: Teachers College Press, Columbia University, 1964.

Durr, W. K., et al. The Houghton-Mifflin readers. Boston: Houghton-Mifflin Co., 1971-1974.

Durr, W. K., et al. The Houghton-Mifflin reading series. Boston: Houghton-Mifflin Co., 1974-1978.

Ekwall, E. Locating and correcting reading difficulties. Columbus, Ohio: Charles E. Merrill Co., 1970.

Emans, R. The usefulness of phonic generalizations above the primary level. The Reading Teacher, 1967, 20, 419-425.

Farr, R. C., Prescott, G. A., Balow, I. H., & Hogan, T. P. Metropolitan achievement tests, Form JI. New York: Harcourt Brace Jovanovich, Inc., 1978.

Farr, R., & Roser, N. L. Reading assessment: A look at problems and issues. Journal of Reading, 1974, 17(4), 592-599.

Farr, R., & Roser, N. L. Teaching a child to read. New York: Harcourt Brace Jovanovich, Inc., 1979.

Fox, B., & Routh, D. K. Phonemic analysis and synthesis as word-attack skills. Journal of Educational Psychology, 1976, 68, 70-74.

Freebery, J. W. Reading achievement down, discipline problems up: Reversing this middle school trend. Individual Practicum for Dissertation, Nova University, Fort Lauderdale, Florida, 1978. [ERIC Document Reproduction Service No. ED 154 371]

Garrison, S. C., & Heard, M. T. An experimental study of the value of phonetics. Peabody Journal of Education, 1931, 9, 9-14.

Gates, A. I. The improvement of reading (3rd ed.). New York: Macmillan, 1947.

Glass, G. G. The teaching of word analysis through perceptual conditioning. Reading and Inquiry, IRA Proceedings, 1965, 10.

Glaser, R. Instructional technology and the measurement of learning outcomes. American Psychologist, 1963, 18, 519-521.

Glickman, C. D. Mastery learning stifles individuality. Journal of Educational Leadership, 1979, 37(2), 100-101.

Goodman, K., Farr, R., & Cassidy, J. Minimum competency standards: Three points of view. Newark, Del.: International Reading Association, 1978.

Gray, W. S. On their own in reading. Chicago, Ill.: Scott, Foresman, 1960.

Groff, P. The syllable: Its nature and pedagogical usefulness. Portland, Ore.: Northwest Regional Educational Laboratory, 1971.

Groff, P. The new anti-phonics. Elementary School Journal, 1977, 77(4), 323.

Guthrie, J. T. Models of reading and reading disability. Journal of Educational Psychology, 1973, 65(1), 9-18.

Hambleton, R., Swaminathan, H., Algina, J., & Coulson, D. Criterion-referenced testing and measurement: A review of technical issues and developments. Review of Educational Research, 1978, 48(1), 1-47.

Hannon, J. P. The Chicago plan: Mastery learning in the Chicago public schools. Educational Leadership, 1979, 37(2), 120-122.

Huynh, H. Statistical consideration of mastery scores. Psychometrica, 1976, 41, 65-78.

Hyman, J. S., & Cohen, S. A. Learning for mastery: Ten conclusions after 15 years and 3,000 schools. Journal of Educational Leadership, 1979, 37(2), 104-109.

Jeffrey, W. E., & Samuels, S. J. The effect of method of reading training on initial learning and transfer. Journal of Verbal Learning and Verbal Behavior, 1967, 6, 354-358.

Jenkins, J. R., Bausell, R. B., & Jenkins, L. M. Comparison of letter name and letter sound training as transfer variables. American Educational Reserach Journal, 1972, 9, 75-86.

Johnson, D. D., & Buamann, J. F. Ginn Lexicon project: Final report. Lexington, Mass.: Ginn & Company, 1979.

Johnson, D. D., & Baumann, J. F. Word identification--instructional practices: The state of the art. In P. D. Pearson (Ed.), Handbook of research in reading. New York: Longman Publishers, in press.

Johnson, D. D., & Pearson, P. D. Teaching reading vocabulary. New York: Holt, Rinehart, & Winston, 1978.

Johnson, D. D., Pittelman, S. D., Schwenker, J., & Shriberg, L. K. The assessment of structural analysis skills (Technical Report No. 526). Madison: Wisconsin Research and Development Center for Individualized Schooling, 1979.

Johnson, D. D., Pittelman, S. D., Schwenker, J., & Shriberg, L. K. Interim report: The refinement of the test battery to assess word identification skills (Technical Report No. 544). Madison: Wisconsin Research and Development Center for Individualized Schooling, 1980.

Johnson, D. D., Pittelman, S. D., Schwenker, J., Shriberg, L. K., & Morgan-Janty, C. Relationships between word identification skills and reading comprehension of elementary school children (Technical Report No. 490). Madison: Wisconsin Research and Development Center for Individualized Schooling, 1978.

Johnson, D. D., Shriberg, L. K., Pittelman, S. D., & Schwenker, J. A new approach to the assessment of phonics skills (Technical Report No. 532). Madison: Wisconsin Research and Development Center for Individualized Schooling, 1979.

Johnson, M. S., et al. The read system (teacher's annotated ed.). New York: American Book Co., 1968-1972.

Karlin, R. Teaching elementary reading: Principles and strategies. New York: Harcourt Brace Jovanovich, 1971.

Katims, M. Unsystematic observations on success. Journal of Educational Leadership, 1979, 37(2), 118-123.

Katims, M., Smith, J. K., Steel, C., & Wick, J. W. The Chicago mastery learning reading program: An interim evaluation. Paper presented at the annual meeting of the American Educational Research Association, New York City, 1977.

Kean, J., & Personke, C. The language arts: Teaching and learning in the elementary school. New York: St. Martin's Press, 1976.

Keller, F. S. Goodbye, teacher... . Journal of Applied Behavioral Analysis, 1968, 1, 79-89.

Kenyon, J. S., & Knott, T. A. A pronouncing dictionary of American English. Springfield, Mass.: G. & C. Merriam Co., 1953.

Koenke, K. Testing the minimal competencies. The Reading Teacher, 1979, 23(1), 118-122.

Kress, R. A. Review of the Sipay Word Analysis Test. In O. K. Buros (Ed.), The seventh mental measurements yearbook (Vol. II). Highland Park, N.J.: Gryphon Press, 1972, 1292-1293.

Kriewall, T. Aspects and applications of criterion referenced tests. Downers Grove, Ill.: Institute for Educational Research, 1972. [ERIC Document Reproduction Service No. ED 063 333]

Kulik, J., Kulik, C., & Cohen, P. A meta-analysis of outcome studies of Keller's personalized system of instruction. American Psychologist, 1979, 34, 307-318.

Lamb, P., & Arnold, R. Reading foundations and instructional strategies. Belmont, Calif.: Wadsworth Publishing Company, Inc., 1976.

Levine, H. G., & Forman, P. M. A study of retention of knowledge of neurosciences information. Journal of Medical Education, 1973, 48, 867-869.

Lloyd, D. J., & Warfel, H. R. Punctuation. In P. C. Burns, J. E. Alexander, & A. R. Davis (Eds.), Language arts concepts for elementary school teachers. Itasca, Ill.: F. E. Peacock, 1972.

Madden, R., Gardner, E. F., Rudman, H. C., Karlsen, B., & Merwin, J. C. Stanford achievement test. New York: Harcourt Brace Jovanovich, Inc., 1970-1974.

Massaro, D. W. Understanding language: An information processing analysis of speech perception, reading, and psycholinguistics. New York: Academic Press, 1975.

McDonald, T. F. Pro: Minimal Competency testing as viewed from the front line. Journal of Reading, 1978, 22(1), 13-15.

Meskauskas, J. A. Evaluation models for criterion-referenced testing: Views regarding mastery and standard setting. Review of Educational Research, 1976, 46(1), 133-158.

Meskauskas, J. A., & Webster, G. W. The American Board of Internal Medicine recertification examination process and results. Annals of Internal Medicine, 1975, 82, 577-581.

Miller, G. A., & Nicely, P. E. An analysis of perceptual confusions among some English consonants. Journal of the Acoustical Society of America, 1955, 27, 338-352.

Millman, J. Passing scores and test lengths for domain-referenced measures. Review of Educational Research, 1973, 46, 205-216. [ERIC Document Reproduction Service ED 065 555]

Moyle, D. Sequence and structure in reading development. In J. Merritt (Ed.), New horizons in reading. Newark, Del.: International Reading Association, 1974.

Muller, D. Phonic blending and transfer of letter training to word reading in children. Journal of Reading Behavior, 1973, 5(3), 13-15.

National Assessment of Educational Progress. Writing mechanics, 1969-1974: A capsule description of changes in writing mechanics. Washington, D.C.: National Center for Educational Statistics, 1975. [ERIC Document Reproduction Service No. ED 113 736]

Osburn, W. J. Teaching spelling by teaching syllables and root words. Elementary School Journal, 1954, 55(1), 32-41.

Osburn, W. L. Remedial and follow-up vocabulary (Bull. No. 1). Bloomington, Ill.: Public School Publishing Company, 1925.

Otto, W., & Chester, R. Objective-based reading. Reading, Mass: Addison-Wesley, 1976.

Otto, W., Miles, P. J., Kamm, K., & Stewart, D. M. The Wisconsin design for reading skill development. Minneapolis, Minn.: National Computer Systems, Inc., 1972-1975.

Pearson, P. D., & Johnson, D. D. Teaching reading comprehension. New York: Holt, Rinehart, & Winston, 1978.

Peterson, G. E., & Barney, H. L. Control methods used in a study of the vowels. Journal of the Acoustical Society of America, 1952, 24, 175-184.

Pikulski, J. J., & Shanahan, T. A comparison of various approaches to evaluating phonics. The Reading Teacher, 1980, 33, 692-702.

Poggio, J. Long-term cognitive retention resulting from the mastery learning paradigm. Paper presented to the American Educational Research Association, San Francisco, 1976.

Popham, J. W. Criterion-referenced measurement. Englewood Cliffs, N.J.: Prentice-Hall, 1978.

Popham, J. W. Educational measurement for the improvement of instruction. Phi Delta Kappan, 1980, April, 531-534.

Popham, J. W., & Baker, E. Establishing instructional goals. Englewood Cliffs, N.J.: Prentice-Hall, 1970.

Prescriptive reading inventory. Monterey, Calif.: CTB/McGraw-Hill, 1972 and 1976.

Purves, A. Minimal competencies and measures of competence. SLATE (Support Learning and Teaching of English), 1976, 2(2), 1.

Rhode, M., & Cronnell, B. Compilation of a communication skills lexicon coded with linguistic information (Technical Report No. 58). Los Alamitos, Calif.: SWRL Educational Research and Development, 1977.

Robinson, H., Monroe, M., Artley, A., Huck, C., & Jenkins, W. More roads to follow. Glenview, Ill.: Scott Foresman & Co., 1965.

Romberg, T. A., Shepler, J., & King, I. Mastery learning and retention (Technical Report No. 151). Madison: Wisconsin Research and Development Center for Cognitive Learning, 1970.

Rupley, W., & Longnion, B. L. Mastery learning: A viable alternative? The Reading Teacher, 1978, 32(3), 380-383.

Samuels, S. J. Hierarchical subskills in the reading acquisition process. In J. Guthrie (Ed.), Aspects of reading acquisition. Baltimore, Md.: Johns Hopkins Press, 1976.

Schubert, D. Dictionary of terms and concepts in reading. Springfield, Ill.: Charles C. Thomas Publishers, 1969.

Seymour, D. Z. Let's reconsider minimum competence standards. Journal of Reading, 1979, 23(3), 215-221.

Shepherd, J. F. Morpheme knowledge test. Reviewed in W. T. Fagan, C. R. Cooper, & J. M. Jensen, Measures for research and evaluation in the English language arts. Urbana, Ill.: NCTE Committee on Research, 1973. [ERIC Document Reproduction Service No. ED 091 730]

Sipay, E. R. Sipay word analysis tests. Cambridge, Mass.: Educators Publishing Service, 1974.

Skills monitoring system for reading and word identification. New York: Harcourt Brace Jovanovich, Inc., 1975.

Smith, C. B., & Wardhaugh, R. The new Macmillan reading program: Series r. New York: Macmillan, 1975.

Smith, J. K., & Wick, J. W. Practical problems of attempting to implement a mastery learning program in a large city school. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, 1976.

Smith, N. B. Reading instruction for today's child. Englewood Cliffs, N.J.: Prentice-Hall, 1963.

Smith, R. J., & Johnson, D. D. Teaching children to read. Reading, Mass.: Addison-Wesley Publishing Company, 1976.

Spache, G. D. Toward better reading. Champaign, Ill.: Garrard Publishing Company, 1963.

Spache, G. D. Diagnosing and correcting reading disabilities. Boston, Mass.: Allyn and Bacon, 1976.

Spache, G. D., & Baggett, M. What do teachers know about phonics and syllabication? The Reading Teacher, 1966, 19, 96-99.

Spady, W. G. Competency based education: A bandwagon in search of a definition. Educational Researcher, 1977, 6(2), 9-14.

Stallard, C. Comparing objective based reading programs. Journal of Reading, 1977, 21(1), 36-44. (a)

Stallard, C. Managing reading instruction: Comparative analysis of objective-based reading programs. Educational Technology, 1977, 17(12), 21-26. (b)

Stauffer, R. G. A study of prefixes in the Thorndike list to establish a list of prefixes that should be taught in the elementary school. Journal of Educational Research, 1942, 35(6), 453-458.

Stauffer, R. G. Directing reading maturity as a cognitive process. New York: Harper & Row, 1969.

Swaminathan, R. J., Block, J. H., & Ravitch, M. M. Criterion-referenced testing: Comments on reliability. Journal of Educational Measurement, 1972, 9, 133-137.

Tate, H. L. The influence of phonics on silent reading in grade 1. Elementary School Journal, 1937, 37, 752-763.

Terwilliger, J. S. Some problems associated with the concept of mastery. Unpublished manuscript, University of Minnesota, March 1970.

Terwilliger, J. S. An examination of the concept of mastery. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, April 1979.

Tierney, R. C. Con: Minimal competency testing won't do it. Journal of Reading, 1978, 22(1), 10-12.

Torshen, K. P. The mastery approach to competency based education. New York: Academic Press, 1977.

Valentine, C. W. Experiments on the methods of teaching reading. Journal of Experimental Pedagogy, 1913, 2, 99-112.

Van Roekel, B. H., & Kluwe, M. J. The Harper and Row basic reading program. Evanston, Ill.: Harper & Row Publishers, 1966.

Venezky, R. L. English orthography: Its graphical structure and its relation to sound. Reading Research Quarterly, 1967, 2, 75-106.

Venezky, R. L. The structure of English orthography. The Hague: Mouton, 1970.

Venezky, R. L., & Massaro, D. W. The role of orthographic regularity in word recognition. In L. B. Resnick & P. A. Weaver (Eds.), Theory and practice of early reading (Vol. 1). Hillsdale, N.J.: Lawrence Erlbaum Associates, 1979, 85-107.

Wardhaugh, R. Syl-lab-i-ca-tion. Elementary English, 1966, 63, 785-788.

Westbury, I. Curriculum evaluation. Review of Educational Research, 1970, 40(2), 239-260.

Witty, P. A., Bebell, M. H., & Freeland, A. M. Reading caravan (Rev. ed.). Boston: D.C. Heath & Co., 1968.

Witty, P., Freeland, A., & Grotberg, E. The teaching of reading. Boston: D.C. Heath & Co., 1966.

Zuck, L. V. Some questions about the teaching of syllabication rules. The Reading Teacher, 1974, 27, 583-588.

Bibliography

Abrams, J. D. Mastery learning in a smaller school system. Educational Leadership, 1979, 37(2), 136-139.

Algra, C. Meeting the challenge of a minimum reading graduation requirement. Journal of Reading, 1978, 21(5), 392-397.

Anderson, L. W. Adaptive education. Educational Leadership, 1979, 37(2), 140-143.

Andrews, B. J., & Hecht, J. T. A preliminary investigation of two procedures for setting examination standards. Educational and Psychological Measurement, 1976, 36, 45-50.

Barone, T. E. Reading, writing, and mastery learning: Are they compatible? Educational Leadership, 1978, 36, 187-189.

Block, J. H. The effects of various levels of performance on selected cognitive, affective, and time variables. Unpublished doctoral dissertation, University of Chicago, 1970.

Block, J. H. Individualized instruction: A mastery learning perspective. Educational Leadership, 1977, 34(5), 337-341.

Block, J. H., & Anderson, L. Mastery learning in classroom instruction. New York: Macmillan Publishing Co., 1975.

Block, J. H., & Burns, R. B. Mastery learning. In L. S. Shulman (Ed.), Review of research in education (Vol. 4). Itasca, Illinois: F. E. Peacock Publishers, Inc. and the American Educational Research Association, 1976.

Cooke, L. M. Why business supports mastery learning. Educational Leadership, 1979, 37(2), 124-125.

Cronbach, L. J. Book review of J. H. Block, editor. Mastery learning: Theory and practice. International Review of Education, 1972, 18, 250-252.

Denton, J. J., & Henson, K. T. Mastery learning and grade inflation. Educational Leadership, 1979, 37(2), 150-152.

Ebel, R. L. Essentials of educational measurement. Englewood Cliffs, N.J.: Prentice Hall, 1972.

Farr, R., & Olshavsky, J. E. Is minimum competency testing the appropriate solution to the SAT decline? Phi Delta Kappan, 1980, 61(8), 528-530.

Fry, E. Elementary reading instruction. New York: McGraw-Hill Book Company, Inc., 1977.

Good, C. G. Organizing reading management systems in urban districts. The Reading Teacher, 1980, 33(7), 816-818.

Gusert, G. SCIP: A New Orleans solution to a national problem. Educational Leadership, 1979, 37(2), 128-134.

Hambleton, R. K. Test score validity and standard-setting methods. In R. A. Berk (Ed.), Criterion-referenced measurement: The state of the art. Baltimore, Maryland: The John Hopkins University Press, 1980.

Horton, L. Mastery learning: Sound in theory but... Educational Leadership, 1979, 37(2), 154-156.

Johnson, D. D. Skills management systems: Some issues. Language Arts, 1977, 54(5), 510-516.

Jolly, T. Mastery learning and reading. Journal of Reading, 1979, 22(8), 764-767.

Mather, P. W. Skill management systems in reading: A real revelation. Journal of Reading, 1977, 21(2), 139-144.

McNeill, J. D. Forces influencing curriculum. Review of Educational Research, 1969, 39, 293-318.

Millman, J. Criterion-referenced measurement. In W. J. Popham (Ed.), Evaluation in education: Current practices. Berkeley, Calif.: McCutchan Publishers, 1974.

Montieth, M. K. Minimum competencies, minimal competency testing: Reading and language arts. Journal of Reading, 1978, 21(8), 750-752.

Moore, J. T. A systems approach to individualized reading. The Reading Teacher, 1979, 32(8), 951-955.

Mueller, D. J. Mastery learning: Partly boon, partly boondoggle. Teacher's College Record, September 1976, pp. 41-52.

Nedelsky, L. Absolute grading standards for objective tests. Educational and Psychological Measurement, 1954, 14, 3-19.

Olson, M. A. Criterion-referenced reading assessment in a large city school district. Journal of Reading Behavior, 1976, 8(4), 387-395.

Popham, W. J. Normative data for criterion-referenced tests? Phi Delta Kappan, 1976, 57(9), 593-594.

Rude, R. T. Objective-based reading systems: An evaluation. The Reading Teacher, November 1974, pp. 169-175.

Shavelson, R. J., Block, J. H., & Ravitch, M. M. Criterion-referenced testing: Comments on reliability. Journal of Educational Measurement, 1972, 9, 133-137.

Sheehan, D. S., & Marcus, M. Validating criterion-referenced reading tests, Journal of Reading Behavior, 1977, 9(2), 129-135.

Smith, J. K. Perspectives on mastery learning and mastery testing (ERIC IIM Report 63). Princeton, N.J.: ERIC Clearinghouse on Tests, Measurement, and Evaluation, Educational Testing Service, 1977. [ED 152 813]

Smith, J. K., & Katims, M. Reading in the city: The Chicago mastery learning reading program. Phi Delta Kappan, November 1977, pp. 199-202.

Spiegal, D. L. Holistic approach to diagnosis and remediation. The Reading Teacher, 1974, 27(4), 370-374.

Terwilliger, J. S. Quotas, task difficulty and performance standards for criterion-referenced tests. A paper presented at the annual meeting of the National Council on Measurement in Education, Washington, D.C., April 2, 1975.

Thompson, R. A., & Dziuban, C. D. Criterion-referenced reading tests in perspective. The Reading Teacher, December 1973, pp. 292-294.

Trogdon, W. E. An exercise in mastery learning. Phi Delta Kappan, 1980, 61(6), 389-391.

Center Planning and Policy Committee

Marshall S. Smith
Director

Dale D. Johnson
Area Chairperson
Studies in Language:
Reading and Communication

Marvin J. Fruth
Area Chairperson
Studies in Implementation
of Individualized Schooling

Penelope L. Peterson
Area Chairperson
Studies of Instructional Programming
for the Individual Student

James M. Lipham
Area Chairperson
Studies of Administration and
Organization for Instruction

Thomas A. Romberg
Area Chairperson
Studies in Mathematics and Evaluation
of Practices in Individualized Schooling

Associated Faculty

Vernon L. Allen
Professor
Psychology

B. Dean Bowles
Professor
Educational Administration

Thomas P. Carpenter
Professor
Curriculum and Instruction

W. Patrick Dickson
Assistant Professor
Child and Family Studies

Lloyd E. Frohreich
Professor
Educational Administration

Marvin J. Fruth
Professor
Educational Administration

Dale D. Johnson
Professor
Curriculum and Instruction

Herbert J. Klausmeier
V.A.C. Henmon Professor
Educational Psychology

Joel R. Levin
Professor
Educational Psychology

James M. Lipham
Professor
Educational Administration

Dominic W. Massaro
Professor
Psychology

Donald M. McIsaac
Professor
Educational Administration

Wayne Otto
Professor
Curriculum and Instruction

Penelope L. Peterson
Associate Professor
Educational Psychology

Thomas S. Popkewitz
Professor
Curriculum and Instruction

Gary G. Price
Assistant Professor
Curriculum and Instruction

W. Charles Read
Professor
English and Linguistics

Thomas A. Romberg
Professor
Curriculum and Instruction

Richard A. Rossmiller
Professor
Educational Administration

Peter A. Schreiber
Associate Professor
English and Linguistics

B. Robert Tabachnick
Professor
Curriculum and Instruction

Gary G. Wehlage
Professor
Curriculum and Instruction

Louise Cherry Wilkinson
Associate Professor
Educational Psychology

Steven R. Yussen
Professor
Educational Psychology

10/80